Guang-Zhong Yang David Hawkes
Daniel Rueckert Alison Noble
Chris Taylor (Eds.)

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009

LNCS 5761

12th International Conference
London, UK, September 2009
Proceedings, Part I

MICCAI

Springer

Lecture Notes in Computer Science 5761

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Guang-Zhong Yang David Hawkes
Daniel Rueckert Alison Noble
Chris Taylor (Eds.)

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009

12th International Conference
London, UK, September 20-24, 2009
Proceedings, Part I

Springer

Volume Editors

Guang-Zhong Yang
Imperial College London, Institute of Biomedical Engineering
London, UK
E-mail: g.z.yang@imperial.ac.uk

David Hawkes
University College London, Centre for Medical Image Computing
London, UK
E-mail: d.hawkes@ucl.ac.uk

Daniel Rueckert
Imperial College London, Department of Computing
London, UK
E-mail: d.rueckert@imperial.ac.uk

Alison Noble
University of Oxford, Institute of Biomedical Engineering
Oxford, UK
E-mail: noble@robots.ox.ac.uk

Chris Taylor
University of Manchester, School of Computer Science
Manchester, UK
E-mail: chris.taylor@manchester.ac.uk

Library of Congress Control Number: 2009934167

CR Subject Classification (1998): I.5, I.2.10, I.2.9, J.3, J.6, I.4, K.4.1

LNCS Sublibrary: SL 6 – Image Processing, Computer Vision, Pattern Recognition, and Graphics

ISSN 0302-9743
ISBN-10 3-642-04267-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-04267-6 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 12753628 06/3180 5 4 3 2 1 0

Preface

The 12th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2009, was held in London, England at Imperial College during September 20–24, 2009. The venue was situated in one of London's finest locations, adjacent to landmarks such as The Royal Albert Hall and the Science, Natural History and Victoria and Albert Museums, with Hyde Park just a short walk away.

Over the last decade, the MICCAI conferences have become a premier international event, with papers of very high standard addressing the multidisciplinary fields of biomedical image computing, computer-assisted intervention and medical robotics. The conference has attracted annually leading scientists, engineers and clinicians from a wide range of disciplines.

This year, we received a record submission of 804 papers from 36 different countries worldwide. These covered medical image computing (functional and diffusion image analysis, segmentation, physical and functional modelling, shape analysis, atlases and statistical models, registration, data fusion and multiscale analysis), computer-assisted interventions and robotics (planning and image guidance of interventions, simulation and training systems, clinical platforms, visualization and feedback, robotics and human–robot interaction), and clinical imaging and biomarkers (computer-aided diagnosis, organ/system specific applications, molecular and optical imaging and imaging biomarkers).

A careful, systematic review process was put in place to ensure the best possible program for MICCAI 2009. The Program Committee (PC) of the conference was composed of 39 members, each with recognized international reputation in the main topics covered by the conference. Each one of the 804 submitted papers was assigned to two PC members (a Primary and a Secondary). At least three external reviewers (outside the PC) were assigned to each paper according to their expertise. These external reviewers provided double-blind reviews of the papers, including those submitted by the conference organizers. All reviewers, except a handful who provided last minute "emergency" reviews, refereed between 8 and 10 papers each, giving each reviewer a reasonable sample for ranking the relative quality of the papers. Authors were given the opportunity to rebut the anonymous reviews.

Then, each PC member graded (typically 20) papers as primary based on the external reviews, the rebuttal and his/her own reading of the papers. In addition he/she provided input, as Secondary PC, to typically 20 more papers assigned to various Primary PCs. In summary, each paper was graded by two PC members and three external reviewers (i.e., by five assessors). During a two-day PC meeting involving the PC members held during May 17–18, 2009, papers were selected in a three-stage process:

- First stage: initial acceptance of those papers ranked very high and rejection of those papers ranked very low. Eight groups were formed, each comprising four or five PC members. The groups considered acceptance of the top three papers from each PC member and rejection of the bottom eight. Any papers in doubt were transferred to the second stage.
- Second stage: the same groups of PC members ranked the remaining papers and accepted between 9 and 18 of the highest ranking papers per group and rejected between 18 and 32 of the lowest ranking papers.
- Third stage: a different set of groups were formed and assigned the remaining undecided papers to the "accept" or "reject" category through an iterative process.

In all, we accepted 259 papers (32%) to be included in the proceedings of MICCAI 2009. Of these, 43 were selected for podium presentation (5%) and 216 for poster presentation at the conference (27%).

The review process was developed from that used in previous MICCAI conferences. In particular we are grateful to Rasmus Larsen for his input on the statistical basis for the protocol. Each step of the process ensured that, for random selections of papers to PC members, the probability of correctly assigning rejections and acceptances was at least 95%. With the combined skill and expertise of the PC, we are confident that it exceeded this figure and that we ran a robust system. Acceptance of papers at MICCAI is a competitive process and with such a strong submission rate it is inevitable that many good papers were not able to be included in the final program and we understand the frustration of authors. We too have had many papers rejected. We congratulate those who had papers accepted and encourage those who did not to persevere and submit again next year.

We wish to thank the reviewers and the PC for giving up their precious time ensuring the high quality of reviews and paper selection. These tasks are time consuming and require skill and good judgment, representing a significant effort by all. The continued improvement in the quality of the conference is entirely dependent on this tremendous effort.

We particularly wish to thank James Stewart of *precisionconference.com* for the efficient organization of the website and rapid response to any queries and requests for changes, many of them totally unreasonable and at a very short notice.

One highlight of MICCAI 2009 was the workshops and tutorials organized before and after the main conference. We had a record number of submissions which resulted in a very exciting, diverse and high-quality program. The workshops provided a comprehensive coverage on topics that were not fully explored during the main conference, including "grand challenges," and some emerging areas of MICCAI, whereas the tutorials provided educational material for training new professionals in the field including students, clinicians and new researchers. We are grateful to all workshop and tutorial organizers for making these events a great success.

We would also like to thank our two invited keynote speakers, Sir Michael Brady, University of Oxford, UK, and Koji Ikuta, Nagoya University, Japan. Their presentations on "Oncological Image Analysis" and "Nano and Micro Robotics for Future Biomedicine" were both inspiring and entertaining.

The conference would not be possible without the commitment and hard work of the local organizing team. In particular, we thank our Associate Editors Adrian Chung and Su-Lin Lee for their help in working with all authors in improving the final manuscript, and Dominique Drai, Ron Gaston, Thomy Merzanidou, Christiana Christodoulou, Karim Lekadir, Felipe Orihuela-Espina, Lichao Wang, Fani Deligianni, and Dong Ping Zhang for checking the original submissions and for assisting in the compilation of the proceedings.

We are grateful to Ferdinando Rodriguez y Baena for coordinating the corporate sponsorship and industrial/academic exhibitions, Dan Elson and Fernando Bello for coordinating MICCAI workshops and tutorials, Eddie Edwards for managing the conference registration and social events, and Raphaele Raupp for assisting with all the conference logistics. We also thank Robert Merrifield for his kind help in graphics design and George Mylonas for his huge effort in designing and implementing the hardware/software platforms for the conference e-Teaser sessions.

We are extremely grateful to Betty Yue, Ulrika Wernmark and their team for their tireless effort in managing all aspects of the conference organization—it is through their effort that we managed to have a seamless event on a busy campus where many dedicated facilities including the fully equipped poster hall had to be installed specially for the conference. We also thank all the session Chairs and Co-chairs in managing and coordinating the presentations during the conference.

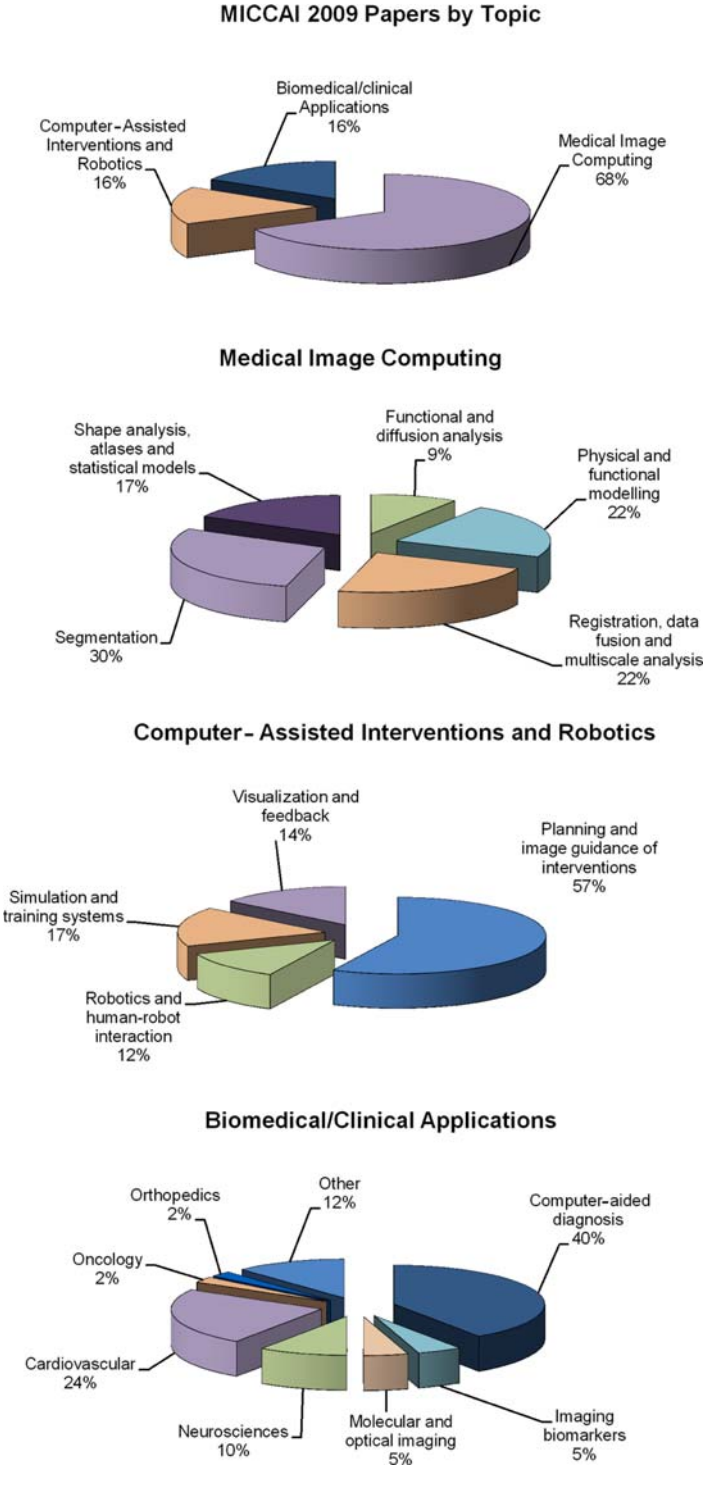
We would also like to thank the MICCAI Society for providing valuable input and support to the conference, especially Guido Gerig for coordinating the MICCAI Young Scientist Awards and Richard Robb for coordinating the Best Paper Awards.

Last but not least, we would like to thank all our sponsors for their kind support, particularly in this most difficult economic climate. Their generosity ensured the highest quality of the conference and essential support to students and young researchers.

It was our pleasure to welcome the MICCAI 2009 attendees to London. In addition to attending the workshop, we trust that the attendees also took the opportunity to explore the rich culture and history of the city. We look forward to meeting you again at MICCAI 2010 in Beijing, China.

September 2009

Guang-Zhong Yang
David Hawkes
Daniel Rueckert
Alison Noble
Chris Taylor


MICCAI 2009 Papers by Topic
Biomedical/clinical Applications 16%
Computer-Assisted Interventions and Robotics 16%
Medical Image Computing 68%
Medical Image Computing
Shape analysis, atlases and statistical models 17%
Functional and diffusion analysis 9%
Physical and functional modelling 22%
Segmentation 30%
Registration, data fusion and multiscale analysis 22%
Computer-Assisted Interventions and Robotics
Visualization and feedback 14%
Planning and image guidance of interventions 57%
Simulation and training systems 17%
Robotics and human-robot interaction 12%
Biomedical/Clinical Applications
Orthopedics 2%
Other 12%
Computer-aided diagnosis 40%
Oncology 2%
Cardiovascular 24%
Neurosciences 10%
Molecular and optical imaging 5%
Imaging biomarkers 5%

Organization

General Chairs

Guang-Zhong Yang	Imperial College London, UK

Program Co-chairs

David Hawkes	University College London, UK
Daniel Rueckert	Imperial College London, UK
Guang-Zhong Yang	Imperial College London, UK

Workshop Chair and Co-chairs

Alison Noble (Chair)	University of Oxford, UK
Nobuhiko Hata	Harvard University, USA
Xavier Pennec	INRIA, France

Tutorial Chair and Co-chairs

Chris Taylor (Chair)	University of Manchester, UK
Hongen Liao	University of Tokyo, Japan
Frederik Maes	KU Leuven, Belgium

Associate Editors

Adrian Chung	Imperial College London, UK
Su-Lin Lee	Imperial College London, UK

Coordinator for MICCAI Young Scientist Awards

Guido Gerig	University of Utah, USA

Coordinator for Best Paper Awards

Richard Robb	Mayo Clinic College of Medicine Rochester, Minnesota, USA

Program Committee

Christian Barillot	IRISA, France
Wolfgang Birkfellner	University of Vienna, Austria
Ela Claridge	University of Birmingham, UK
Tim Cootes	University of Manchester, UK
Randy Ellis	Queen's University, Canada
Gabor Fichtinger	Queen's University, Canada
Jim Gee	University of Pennsylvania, Philadelphia, USA
Guido Gerig	University of Utah, USA
Polina Golland	Massachusetts Institute of Technology, USA
Tianzi Jiang	Chinese Academy of Sciences, China
Leo Joskowicz	Hebrew University of Jerusalem, Israel
Rasmus Larsen	Technical University of Denmark, Denmark
Cristian Lorenz	Philips Research Labs, Germany
Frederik Maes	University of Leuven, Belgium
Kensaku Mori	Nagoya University, Japan
Nassir Navab	Technical University of Munich, Germany
Mads Nielsen	University of Copenhagen, Denmark
Poul Nielsen	University of Auckland, New Zealand
Wiro Niessen	Erasmus Medical Center, The Netherlands
Sebastien Ourselin	University College London, UK
Xavier Pennec	INRIA, France
Graeme Penney	King's College London, UK
Franjo Pernus	University of Ljubljana, Slovenia
Terry Peters	Robarts Research Institute, Canada
Ichiro Sakuma	University of Tokyo, Japan
Tim Salcudean	University of British Columbia, Canada
Yoshinobu Sato	University of Osaka, Japan
Julia Schnabel	University of Oxford, UK
Dinggang Shen	University of North Carolina, USA
Gabor Szekely	ETH Zurich, Switzerland
Russell Taylor	John Hopkins University, USA
Jocelyne Troccaz	CNRS, France
Max Viergever	University of Utrecht, The Netherlands
Simon Warfield	Harvard University, USA
Sandy Wells	Harvard University, USA
Carl-Frederik Westin	Harvard University, USA
Chenyang Xu	Siemens Corporate Research, USA

Local Organizing Committee

Ferdinando Rodriguez y Baena	Imperial College London, UK
Fernando Bello	Imperial College London, UK
Brian Davies	Imperial College London, UK
Eddie Edwards	Imperial College London, UK
Daniel Elson	Imperial College London, UK
David Firmin	Imperial College London, UK
Andrew Todd-Pokropek	University College London, UK
Ulrika Wernmark	Imperial College London, UK
Betty Yue	Imperial College London, UK

Reviewers

Abolmaesumi, Purang
Abugharbieh, Rafeef
Alexander, Daniel
Aljabar, Paul
Alvino, Christopher
Amini, Amir
An, Jungha
Angelini, Elsa
Arridge, Simon R.
Ashburner, John
Astley, Sue
Atkinson, David
Awate, Suyash
Axel, Leon
Aylward, Stephen

Babalola, Kolawole
Barratt, Dean
Batchelor, Philip
Baumann, Michael
Bazin, Pierre-Louis
Beckmann, Christian
Beichel, Reinhard
Bello, Fernando
Berger, Marie-Odile
Betke, Margrit
Bhalerao, Abhir
Bhotika, Rahul
Bischof, Horst
Blezek, Daniel
Bloch, Isabelle
Boctor, Emad
Bouix, Sylvain
Boukerroui, Djamal
Bourgeat, Pierrick
Brady, Mike
Bromiley, Paul
Buelow, Thomas
Bullitt, Elizabeth

Camara, Oscar
Camp, Jon
Cattin, Philippe C.
Chakravarty, M. Mallar
Chou, Yiyu
Christensen, Gary
Chung, Adrian
Chung, Albert C. S.
Cinquin, Philippe
Clark, Chris
Colchester, Alan
Collins, D. Louis
Colliot, Olivier
Comaniciu, Dorin
Commowick, Olivier
Cook, Philip
Coulon, Olivier
Crozier, Stuart
Crum, William

Dam, Erik
Darvann, Tron

Dauguet, Julien
Davatzikos, Christos
Dawant, Benoit
de Bruijne, Marleen
Deligianni, Fani
Delingette, Hervé
Deriche, Rachid
Descoteaux, Maxime
Desvignes, Michel
Douiri, Abdel
Duan, Qi
Duchesne, Simon
Duncan, James S
Dupont, Pierre

Ecabert, Olivier
Edwards, Philip
El-Baz, Ayman

Fan, Yong
Farag, Aly
Fenster, Aaron
Feuerstein, Marco
Figl, Michael
Fillard, Pierre
Fischer, Bernd
Fitzpatrick, J Michael
Fleig, Oliver
Fletcher, P. Thomas
Florack, Luc
Frangi, Alejandro
Freysinger, Wolfgang
Funka-Lea, Gareth

Gibaud, Bernard
Gilson, Wesley
Glocker, Ben
Gonzalez Ballester, Miguel Angel
Gooding, Mark
Graham, Jim
Grau, Vicente
Grimson, Eric
Groher, Martin
Guetter, Christoph
Guimond, Alexandre

Hager, Gregory D
Hahn, Horst
Hamarneh, Ghassan
Hanson, Dennis
Harders, Matthias
Hartley, Richard
Hastreiter, Peter
Hata, Nobuhiko
Haynor, David
Heimann, Tobias
Hellier, Pierre
Heng, Pheng Ann
Hipwell, John
Holmes, David
Hornegger, Joachim
Howe, Robert
Hu, Mingxing
Huang, Heng
Huang, Xiaolei
Hunter, Peter

Jain, Ameet
Janke, Andrew
Jannin, Pierre
Jenkinson, Mark
Jolly, Marie-Pierre
Jomier, Julien
Joshi, Sarang

Kabus, Sven
Kakadiaris, Ioannis
Karamalis, Athanasios
Karssemeijer, Nico
Kaus, Michael
Kazanzides, Peter
Keeve, Erwin
Keil, Andreas
Khamene, Ali
Khurd, Parmeshwar
Kikinis, Ron
Kindlmann, Gordon
King, Andrew
Kirchberg, Klaus
Klein, Stefan
Klinder, Tobias

Kruggel, Frithjof
Kurtcuoglu, Vartan
Kutter, Oliver

Laine, Andrew
Langs, Georg
Lapeer, Rudy
Lee, Su-Lin
Lelieveldt, Boudewijn
Lenglet, Christophe
Lepore, Natasha
Li, Chunming
Li, Shuo
Li, Yonghui
Liang, Jianming
Liao, Rui
Likar, Bostjan
Linguraru, Marius George
Liu, Alan
Liu, Tianming
Liu, Yong
Loeckx, Dirk
Loew, Murray
Lohmann, Gabriele
Lötjönen, Jyrki
Lu, Le
Luboz, Vincent

Ma, Burton
Madabhushi, Anant
Makram-Ebeid, Sherif
Malandain, Gregoire
Manduca, Armando
Mangin, Jean-Francois
Manniesing, Rashindra
Marsland, Stephen
Martel, Anne
Martí, Robert
Martin-Fernandez, Marcos
Mattes, Julian
Mazza, Edoardo
McClelland, Jamie
McGraw, Tim
Meijering, Erik
Metaxas, Dimitris
Meyer, Chuck
Miller, James
Milles, Julien
Modersitzki, Jan
Mohamed, Ashraf
Morel, Guillaume
Murgasova, Maria
Mylonas, George

Nain, Delphine
Nash, Martyn
Nicolau, Stephane
Niethammer, Marc
Nishikawa, Atsushi

O'Donnell, Lauren
O'Donnell, Thomas
Ogier, Arnaud
Olivo-Marin, Jean-Christophe
Olszewski, Mark

Pang, Wai-Man
Papademetris, Xenios
Paragios, Nikos
Patriciu, Alexandru
Paul, Perrine
Paulsen, Rasmus
Pauly, Olivier
Peitgen, Heinz-Otto
Peyrat, Jean-Marc
Pham, Dzung
Pichon, Eric
Pitiot, Alain
Pluim, Josien
Pohl, Kilian Maria
Poignet, Philippe
Prager, Richard
Prastawa, Marcel
Prause, Guido
Prima, Sylvain

Qian, Xiaoning

Radeva, Petia
Rajagopal, Vijayaraghavan
Rajagopalan, Srinivasan
Ramamurthi, Krishnakumar

Reinhardt, Joseph
Rexilius, Jan
Reyes, Mauricio
Rhode, Kawal
Ridgway, Gerard
Rittscher, Jens
Riviere, Cameron
Robb, Richard A
Robinson, Emma
Rodriguez y Baena, Ferdinando
Rohde, Gustavo
Rohlfing, Torsten
Rohling, Robert
Rohr, Karl
Rousseau, François
Russakoff, Daniel

Sabuncu, Mert Rory
Salvado, Olivier
Schweikard, Achim
Sermesant, Maxime
Shechter, Guy
Shen, Xilin
Shi, Pengcheng
Shimizu, Akinobu
Shu, Ni
Siddiqi, Kaleem
Siewerdsen, Jeffrey
Sjöstrand, Karl
Sled, John G.
Smith, Stephen
Sporring, Jon
Staib, Lawrence
Stewart, James
Stoyanov, Danail
Studholme, Colin
Styner, Martin
Subramanian, Navneeth
Suetens, Paul
Summers, Ronald
Sundar, Hari
Szczerba, Dominik
Szilagyi, Laszlo

Tanner, Christine
Tao, Xiaodong
Tasdizen, Tolga
Taylor, Zeike
ter Haar Romeny, Bart M.
Thèvenaz, Philippe
Thiran, Jean-Philippe
Thiriet, Marc
Thirion, Bertrand
Todd Pokropek, Andrew
Toews, Matthew
Tosun, Duygu
Tristán-Vega, Antonio
Tsechpenakis, Gavriil
Twining, Carole

Unal, Gozde

van Assen, Hans
van Ginneken, Bram
Van Leemput, Koen
van Walsum, Theo
Vandermeulen, Dirk
Vannier, Michael
Vercauteren, Tom
Vik, Torbjörn
Villard, Pierre-Frederic
Vrooman, Henri

Wachinger, Christian
Wahle, Andreas
Wang, Defeng
Weber, Stefan
Weese, Jürgen
Wein, Wolfgang
West, Jay
Whitaker, Ross
Whitcher, Brandon
Wiemker, Rafael
Wolf, Ivo
Wong, Stephen
Woolrich, Mark

Xue, Zhong

Yan, Pingkun
Yang, Hua
Yaniv, Ziv

Yendiki, Anastasia
Yeo, Boon Thye
Yoo, Terry
Young, Alistair
Yushkevich, Paul

Zhang, Hui
Zhang, Yong
Zheng, Guoyan
Zheng, Yefeng
Zhou, S. Kevin
Zhou, Xiang (Sean)
Zhu, Hongtu
Zikic, Darko
Zollei, Lilla
Zwiggelaar, Reyer

MICCAI Society, Board of Directors

Nicholas Ayache	INRIA, Sophia Antipolis, France
Kevin Cleary	Georgetown University, Washington DC, USA
James Duncan	Yale University, New Haven, Connecticut, USA
Gabor Fichtinger	Queen's University, Kingston, Ontario, Canada
Guido Gerig	University of Utah, Salt Lake City Utah, USA
Polina Golland	Massachusetts Institute of Technology, USA
Tianzi Jiang	Chinese Academy of Sciences, China
Dimitris Metaxas	Rutgers University, Piscataway, New Jersey, USA
Nassir Navab	Technical University of Munich, Germany
Alison Noble	University of Oxford, Oxford, UK
Sebastien Ourselin	University College London UK
Terry Peters	Robarts Research Institute, London, Ontario, Canada
Richard Robb	Mayo Clinic College of Medicine, Minnesota, USA
Ichiro Sakuma	University of Tokyo, Japan
Guang-Zhong Yang	Imperial College London, UK

MICCAI Society, Executive Officers

President, James Duncan
Executive Director, Richard Robb
Secretary, Nicholas Ayache
Treasurer, Terry Peters
Elections Officer, Karl Heinz Hoehne
Awards Officer, Guido Gerig

MICCAI Society, Staff

Society Secretariat, Janette Wallace, Canada
Membership Coordinator, Gabor Szekely, Switzerland
Publication Coordinator, Nobuhiko Hata, USA
Communications Coordinator, Kirby Vosburgh, USA
Industry Relations Coordinator, Tina Kapur, USA

MICCAI Young Scientist Awards, New York City, 2008

Each year MICCAI awards outstanding work written and presented by students. Both oral and poster presentations are eligible for the awards, and the awards are presented to the winners in a public ceremony at the end of the conference. Six MICCAI Young Scientist Awards were presented by the MICCAI Society at MICCAI 2008:

- *The Effect of Automated Marker Detection on In Vivo Volumetric Stent Reconstruction*: Gert Schoonenberg, Pierre Lelong, Raoul Florent, Onno Wink, Bart ter Haar Romeny, Technische Universiteit Eindhoven, The Netherlands, Philips Healthcare Best NL, Philips France, Paris, France.
- *Passive Ventricular Mechanics Modelling Using Cardiac MR Imaging of Structure and Function*: Vicky Y. Wang, Hoi Leng Lam, Daniel B. Ennis, Alistair A. Young, Martyn P. Nash, University of Auckland New Zealand, UCLA, USA.
- *On-the-Fly Motion-Compensated Cone-Beam CT Using an A Priori Motion Model*: Simon Rit, Jochem Wolthaus, Marcel von Herk, Jan-Jakob Sonke, The Netherlands Cancer Institute, The Netherlands.
- *A Constrained Non-Rigid Registration Algorithm for use in Prostate Image-Guided Radiotherapy*: William Greene, Sudhakar Chelikani, Kailas Purushothaman, Zhe Chen, Jonathan Krisely, Lawrence Staib, Xenophon Papademetris, Jim Duncan, Yale University, USA.
- *Fully Bayesian Joint Model for MR Brain Scan Tissue and Subcortical Structure Segmentation*: Benoit Scherrer, Florence Forbes, Catherine Garbay, Michel Dojat, INSERN, CNRS, INRIA, Université Joseph Fourier, Grenoble France.
- *Sparse Approximation of Currents for Statistics on Curves and Surfaces:* Stanley Durrleman, Xavier Pennec, Alain Trouve, Nicholas Ayache, INRIA Sophia-Antipolis, ENS-Cachan, France.

Elsevier MedIA-MICCAI Prize 2008

Two prizes were awarded by Elsevier during MICCAI 2008 to the first authors of two outstanding articles of the special issue of the *Medical Image Analysis* journal (volume 12, issue 5, October 2008) dedicated to the previous MICCAI 2007 conference.

- First prize awarded to Cyril Poupon for the article: *Real-time MR diffusion tensor and Q-ball imaging using Kalman filtering*, Cyril Poupon, Alexis Roche, Jessica Dubois, Jean-François Mangin, Fabrice Poupon, Medical Image Analysis 12(5) (2008), pages 527-534.
- Second prize awarded to Gabor Fichtinger for the article: *Robotic assistance for ultrasound-guided prostate brachytherapy*, Gabor Fichtinger, Jonathan P. Fiene, Christopher W. Kennedy, Gernot Kronreif, Iulian Iordachita, Danny Y. Song, Everette C. Burdette, Peter Kazanzides, Medical Image Analysis 12(5) (2008), pages 535-545.

Table of Contents – Part I

Cardiovascular Image Guided Intervention and Robotics

Optimal Transseptal Puncture Location for Robot-Assisted Left Atrial Catheter Ablation 1
Jagadeesan Jayender, Rajni V. Patel, Gregory F. Michaud, and Nobuhiko Hata

Towards Guidance of Electrophysiological Procedures with Real-Time 3D Intracardiac Echocardiography Fusion to C-arm CT 9
Wolfgang Wein, Estelle Camus, Matthias John, Mamadou Diallo, Christophe Duong, Amin Al-Ahmad, Rebecca Fahrig, Ali Khamene, and Chenyang Xu

Personalized Pulmonary Trunk Modeling for Intervention Planning and Valve Assessment Estimated from CT Data 17
Dime Vitanovski, Razvan Ioan Ionasec, Bogdan Georgescu, Martin Huber, Andrew Mayall Taylor, Joachim Hornegger, and Dorin Comaniciu

Robotic Force Stabilization for Beating Heart Intracardiac Surgery 26
Shelten G. Yuen, Michael C. Yip, Nikolay V. Vasilyev, Douglas P. Perrin, Pedro J. del Nido, and Robert D. Howe

Non-rigid Reconstruction of the Beating Heart Surface for Minimally Invasive Cardiac Surgery 34
Mingxing Hu, Graeme P. Penney, Daniel Rueckert, Philip J. Edwards, Fernando Bello, Roberto Casula, Michael Figl, and David J. Hawkes

Surgical Navigation and Tissue Interaction

3D Meshless Prostate Segmentation and Registration in Image Guided Radiotherapy 43
Ting Chen, Sung Kim, Jinghao Zhou, Dimitris Metaxas, Gunaretnam Rajagopal, and Ning Yue

A Computer Model of Soft Tissue Interaction with a Surgical Aspirator 51
Vincent Mora, Di Jiang, Rupert Brooks, and Sébastien Delorme

Optimal Matching for Prostate Brachytherapy Seed Localization with Dimension Reduction 59
Junghoon Lee, Christian Labat, Ameet K. Jain, Danny Y. Song, Everette C. Burdette, Gabor Fichtinger, and Jerry L. Prince

Prostate Biopsy Assistance System with Gland Deformation Estimation for Enhanced Precision 67
Michael Baumann, Pierre Mozer, Vincent Daanen, and Jocelyne Troccaz

Prediction of the Repair Surface over Cartilage Defects: A Comparison of Three Methods in a Sheep Model 75
Manuela Kunz, Steven Devlin, Ren Hui Gong, Jiro Inoue, Stephen D. Waldman, Mark Hurtig, Purang Abolmaesumi, and James Stewart

Intra-operative Optical Imaging and Endoscopic Navigation

A Coaxial Laser Endoscope with Arbitrary Spots in Endoscopic View for Fetal Surgery 83
Noriaki Yamanaka, Hiromasa Yamashita, Ken Masamune, Hongen Liao, Toshio Chiba, and Takeyoshi Dohi

Toward Video-Based Navigation for Endoscopic Endonasal Skull Base Surgery 91
Daniel Mirota, Hanzi Wang, Russell H. Taylor, Masaru Ishii, and Gregory D. Hager

Correcting Motion Artifacts in Retinal Spectral Domain Optical Coherence Tomography via Image Registration 100
Susanna Ricco, Mei Chen, Hiroshi Ishikawa, Gadi Wollstein, and Joel Schuman

Single Fiber Optical Coherence Tomography Microsurgical Instruments for Computer and Robot-Assisted Retinal Surgery 108
Marcin Balicki, Jae-Ho Han, Iulian Iordachita, Peter Gehlbach, James Handa, Russell Taylor, and Jin Kang

Motion Modelling and Image Formation

Coronary Tree Extraction Using Motion Layer Separation 116
Wei Zhang, Haibin Ling, Simone Prummer, Kevin Shaohua Zhou, Martin Ostermeier, and Dorin Comaniciu

A Fast Alternative to Computational Fluid Dynamics for High Quality Imaging of Blood Flow 124
Robert H.P. McGregor, Dominik Szczerba, Krishnamurthy Muralidhar, and Gábor Székely

Interventional 4-D Motion Estimation and Reconstruction of Cardiac Vasculature without Motion Periodicity Assumption 132
Christopher Rohkohl, Günter Lauritsch, Marcus Prümmer, and Joachim Hornegger

Estimating Continuous 4D Wall Motion of Cerebral Aneurysms from 3D Rotational Angiography 140
Chong Zhang, Mathieu De Craene, Maria-Cruz Villa-Uriol, Jose M. Pozo, Bart H. Bijnens, and Alejandro F. Frangi

NIBART: A New Interval Based Algebraic Reconstruction Technique for Error Quantification of Emission Tomography Images 148
Olivier Strauss, Abdelkabir Lahrech, Agnès Rico, Denis Mariano-Goulart, and Benoît Telle

Image Registration

A Log-Euclidean Polyaffine Registration for Articulated Structures in Medical Images 156
Miguel Ángel Martín-Fernández, Marcos Martín-Fernández, and Carlos Alberola-López

Nonrigid Registration of Myocardial Perfusion MRI Using Pseudo Ground Truth 165
Chao Li and Ying Sun

Parallax-Free Long Bone X-ray Image Stitching 173
Lejing Wang, Joerg Traub, Simon Weidert, Sandro Michael Heining, Ekkehard Euler, and Nassir Navab

Diffusion Tensor Field Registration in the Presence of Uncertainty 181
Mustafa Okan Irfanoglu, Cheng Guan Koay, Sinisa Pajevic, Raghu Machiraju, and Peter J. Basser

Non-rigid Registration of High Angular Resolution Diffusion Images Represented by Gaussian Mixture Fields 190
Guang Cheng, Baba C. Vemuri, Paul R. Carney, and Thomas H. Mareci

Modelling and Segmentation

Toward Real-Time Simulation of Blood-Coil Interaction during Aneurysm Embolization 198
Yiyi Wei, Stéphane Cotin, Le Fang, Jérémie Allard, Chunhong Pan, and Songde Ma

A Dynamical Shape Prior for LV Segmentation from RT3D Echocardiography 206
Yun Zhu, Xenophon Papademetris, Albert J. Sinusas, and James S. Duncan

A Statistical Model of Right Ventricle in Tetralogy of Fallot for Prediction of Remodelling and Therapy Planning 214
Tommaso Mansi, Stanley Durrleman, Boris Bernhardt, Maxime Sermesant, Hervé Delingette, Ingmar Voigt, Philipp Lurz, Andrew M. Taylor, Julie Blanc, Younes Boudjemline, Xavier Pennec, and Nicholas Ayache

Bayesian Maximal Paths for Coronary Artery Segmentation from 3D CT Angiograms 222
David Lesage, Elsa D. Angelini, Isabelle Bloch, and Gareth Funka-Lea

Image Segmentation and Classification

Hierarchical Normalized Cuts: Unsupervised Segmentation of Vascular Biomarkers from Ovarian Cancer Tissue Microarrays 230
Andrew Janowczyk, Sharat Chandran, Rajendra Singh, Dimitra Sasaroli, George Coukos, Michael D. Feldman, and Anant Madabhushi

Nonparametric Intensity Priors for Level Set Segmentation of Low Contrast Structures 239
Sokratis Makrogiannis, Rahul Bhotika, James V. Miller, John Skinner Jr., and Melissa Vass

Improving Pit–Pattern Classification of Endoscopy Images by a Combination of Experts 247
Michael Häfner, Alfred Gangl, Roland Kwitt, Andreas Uhl, Andreas Vécsei, and Friedrich Wrba

Fast Automatic Segmentation of the Esophagus from 3D CT Data Using a Probabilistic Model 255
Johannes Feulner, S. Kevin Zhou, Alexander Cavallaro, Sascha Seifert, Joachim Hornegger, and Dorin Comaniciu

Automatic Segmentation of the Pulmonary Lobes from Fissures, Airways, and Lung Borders: Evaluation of Robustness against Missing Data 263
Eva M. van Rikxoort, Mathias Prokop, Bartjan de Hoop, Max A. Viergever, Josien P.W. Pluim, and Bram van Ginneken

Segmentation and Atlas Based Techniques

Joint Segmentation of Image Ensembles via Latent Atlases 272
Tammy Riklin Raviv, Koen Van Leemput, William M. Wells III, and Polina Golland

Robust Medical Images Segmentation Using Learned Shape and Appearance Models 281
Ayman El-Baz and Georgy Gimel'farb

A Spatio-temporal Atlas of the Human Fetal Brain with Application to Tissue Segmentation 289
Piotr A. Habas, Kio Kim, Francois Rousseau, Orit A. Glenn, A. James Barkovich, and Colin Studholme

Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets 297
Stanley Durrleman, Xavier Pennec, Alain Trouvé, Guido Gerig, and Nicholas Ayache

On the Manifold Structure of the Space of Brain Images 305
Samuel Gerber, Tolga Tasdizen, Sarang Joshi, and Ross Whitaker

Neuroimage Analysis

Gyral Folding Pattern Analysis via Surface Profiling 313
Kaiming Li, Lei Guo, Gang Li, Jingxin Nie, Carlos Faraco, Qun Zhao, Stephen Miller, and Tianming Liu

Constrained Data Decomposition and Regression for Analyzing Healthy Aging from Fiber Tract Diffusion Properties 321
Sylvain Gouttard, Marcel Prastawa, Elizabeth Bullitt, Weili Lin, Casey Goodlett, and Guido Gerig

Two-Compartment Models of the Diffusion MR Signal in Brain White Matter 329
Eleftheria Panagiotaki, Hubert Fonteijn, Bernard Siow, Matt G. Hall, Anthony Price, Mark F. Lythgoe, and Daniel C. Alexander

Multivariate Tensor-Based Brain Anatomical Surface Morphometry via Holomorphic One-Forms 337
Yalin Wang, Tony F. Chan, Arthur W. Toga, and Paul M. Thompson

Local White Matter Geometry Indices from Diffusion Tensor Gradients 345
Peter Savadjiev, Gordon Kindlmann, Sylvain Bouix, Martha E. Shenton, and Carl-Fredrik Westin

Surgical Navigation and Robotics

i-BRUSH: A Gaze-Contingent Virtual Paintbrush for Dense 3D Reconstruction in Robotic Assisted Surgery 353
Marco Visentini-Scarzanella, George P. Mylonas, Danail Stoyanov, and Guang-Zhong Yang

Targeting Accuracy under Model-to-Subject Misalignments in Model-Guided Cardiac Surgery 361
Cristian A. Linte, John Moore, Andrew D. Wiles, Chris Wedlake, and Terry M. Peters

Patient Specific 4D Coronary Models from ECG-gated CTA Data for Intra-operative Dynamic Alignment of CTA with X-ray Images 369
Coert T. Metz, Michiel Schaap, Stefan Klein, Lisan A. Neefjes, Ermanno Capuano, Carl Schultz, Robert Jan van Geuns, Patrick W. Serruys, Theo van Walsum, and Wiro J. Niessen

Towards Interactive Planning of Coil Embolization in Brain Aneurysms 377
Jeremie Dequidt, Christian Duriez, Stephane Cotin, and Erwan Kerrien

Temporal Estimation of the 3d Guide-Wire Position Using 2d X-ray Images 386
Marcel Brückner, Frank Deinzer, and Joachim Denzler

3-D Respiratory Motion Compensation during EP Procedures by Image-Based 3-D Lasso Catheter Model Generation and Tracking 394
Alexander Brost, Rui Liao, Joachim Hornegger, and Norbert Strobel

System Design of a Hand-Held Mobile Robot for Craniotomy 402
Gavin Kane, Georg Eggers, Robert Boesecke, Jörg Raczkowsky, Heinz Wörn, Rüdiger Marmulla, and Joachim Mühling

Dynamic Active Constraints for Hyper-Redundant Flexible Robots 410
Ka-Wai Kwok, George P. Mylonas, Loi Wah Sun, Mirna Lerotic, James Clark, Thanos Athanasiou, Ara Darzi, and Guang-Zhong Yang

Nonmagnetic Rigid and Flexible Outer Sheath with Pneumatic Interlocking Mechanism for Minimally Invasive Surgical Approach 418
Hiromasa Yamashita, Siyang Zuo, Ken Masamune, Hongen Liao, and Takeyoshi Dohi

Data-Derived Models for Segmentation with Application to Surgical Assessment and Training 426
Balakrishnan Varadarajan, Carol Reiley, Henry Lin, Sanjeev Khudanpur, and Gregory Hager

Task versus Subtask Surgical Skill Evaluation of Robotic Minimally Invasive Surgery 435
Carol E. Reiley and Gregory D. Hager

Development of the Ultra-Miniaturized Inertial Measurement Unit WB3 for Objective Skill Analysis and Assessment in Neurosurgery: Preliminary Results 443
Massimiliano Zecca, Salvatore Sessa, Zhuohua Lin, Takashi Suzuki, Tomoya Sasaki, Kazuko Itoh, Hiroshi Iseki, and Atsuo Takanishi

Novel Endoscope System with Plasma Flushing for Off-Pump Cardiac Surgery 451
Ken Masamune, Tetsuya Horiuchi, Masahiro Mizutani, Hiromasa Yamashita, Hiroyuki Tsukihara, Noboru Motomura, Shinichi Takamoto, Hongen Liao, and Takeyoshi Dohi

Endoscopic Orientation Correction 459
Kurt Höller, Jochen Penne, Armin Schneider, Jasper Jahn, Javier Gutiérrez Boronat, Thomas Wittenberg, Hubertus Feußner, and Joachim Hornegger

Time-of-Flight 3-D Endoscopy 467
Jochen Penne, Kurt Höller, Michael Stürmer, Thomas Schrauder, Armin Schneider, Rainer Engelbrecht, Hubertus Feußner, Bernhard Schmauss, and Joachim Hornegger

In Vivo OCT Coronary Imaging Augmented with Stent Reendothelialization Score 475
Florian Dubuisson, Claude Kauffmann, Pascal Motreff, and Laurent Sarry

Optical Biopsy Mapping for Minimally Invasive Cancer Screening 483
Peter Mountney, Stamatia Giannarou, Daniel Elson, and Guang-Zhong Yang

Biopsy Site Re-localisation Based on the Computation of Epipolar Lines from Two Previous Endoscopic Images 491
Baptiste Allain, Mingxing Hu, Laurence B. Lovat, Richard Cook, Sebastien Ourselin, and David Hawkes

Probabilistic Region Matching in Narrow-Band Endoscopy for Targeted Optical Biopsy 499
Selen Atasoy, Ben Glocker, Stamatia Giannarou, Diana Mateus, Alexander Meining, Guang-Zhong Yang, and Nassir Navab

Tracked Regularized Ultrasound Elastography for Targeting Breast Radiotherapy 507
Hassan Rivaz, Pezhman Foroughi, Ioana Fleming, Richard Zellars, Emad Boctor, and Gregory Hager

Image Guidance for Spinal Facet Injections Using Tracked Ultrasound 516
John Moore, Colin Clarke, Daniel Bainbridge, Chris Wedlake, Andrew Wiles, Danielle Pace, and Terry Peters

Cervical Vertebrae Tracking in Video-Fluoroscopy Using the Normalized Gradient Field 524
Rianne Reinartz, Bram Platel, Toon Boselie, Henk van Mameren, Henk van Santbrink, and Bart ter Haar Romeny

Expertise Modeling for Automated Planning of Acetabular Cup in Total Hip Arthroplasty Using Combined Bone and Implant Statistical Atlases 532
Itaru Otomaru, Kazuto Kobayashi, Toshiyuki Okada, Masahiko Nakamoto, Yoshiyuki Kagiyama, Masaki Takao, Nobuhiko Sugano, Yukio Tada, and Yoshinobu Sato

Wide-Angle Intraocular Imaging and Localization 540
Christos Bergeles, Kamran Shamaei, Jake J. Abbott, and Bradley J. Nelson

Inverse C-arm Positioning for Interventional Procedures Using Real-Time Body Part Detection 549
Christian Schaller, Christopher Rohkohl, Jochen Penne, Michael Stürmer, and Joachim Hornegger

A Method to Correct for Brain Shift When Building Electrophysiological Atlases for Deep Brain Stimulation (DBS) Surgery 557
Srivatsan Pallavaram, Benoit M. Dawant, Rui Li, Joseph S. Neimat, Michael S. Remple, Chris Kao, Peter E. Konrad, and Pierre-François D'Haese

Image Registration

Asymmetric Image-Template Registration 565
Mert R. Sabuncu, B.T. Thomas Yeo, Koen Van Leemput, Tom Vercauteren, and Polina Golland

A Demons Algorithm for Image Registration with Locally Adaptive Regularization 574
Nathan D. Cahill, J. Alison Noble, and David J. Hawkes

A Meta Registration Framework for Lesion Matching 582
Sharmishtaa Seshamani, Purnima Rajan, Rajesh Kumar, Hani Girgis, Themos Dassopoulos, Gerard Mullin, and Gregory Hager

Automatic Robust Medical Image Registration Using a New Democratic Vector Optimization Approach with Multiple Measures 590
Matthias Wacker and Frank Deinzer

Task-Optimal Registration Cost Functions 598
B.T. Thomas Yeo, Mert Sabuncu, Polina Golland, and Bruce Fischl

Hybrid Spline-Based Multimodal Registration Using Local Measures for Joint Entropy and Mutual Information 607
Andreas Biesdorf, Stefan Wörz, Hans-Jürgen Kaiser, Christoph Stippich, and Karl Rohr

A Robust Solution to Multi-modal Image Registration by Combining Mutual Information with Multi-scale Derivatives 616
Philip A. Legg, Paul L. Rosin, David Marshall, and James E. Morgan

Multimodal Image Registration by Information Fusion at Feature Level 624
Yang Li and Ragini Verma

Accelerating Feature Based Registration Using the Johnson-Lindenstrauss Lemma 632
Ayelet Akselrod-Ballin, Davi Bock, R. Clay Reid, and Simon K. Warfield

Groupwise Registration and Atlas Construction of 4th-Order Tensor Fields Using the $\mathbb{R}^+$ Riemannian Metric 640
Angelos Barmpoutis and Baba C. Vemuri

Closed-Form Jensen-Renyi Divergence for Mixture of Gaussians and Applications to Group-Wise Shape Registration 648
Fei Wang, Tanveer Syeda-Mahmood, Baba C. Vemuri, David Beymer, and Anand Rangarajan

Attribute Vector Guided Groupwise Registration 656
Qian Wang, Pew-Thian Yap, Guorong Wu, and Dinggang Shen

Statistical Regularization of Deformation Fields for Atlas-Based Segmentation of Bone Scintigraphy Images 664
Karl Sjöstrand, Mattias Ohlsson, and Lars Edenbrandt

Graphical Models and Deformable Diffeomorphic Population Registration Using Global and Local Metrics 672
Aristeidis Sotiras, Nikos Komodakis, Ben Glocker, Jean-François Deux, and Nikos Paragios

Efficient Large Deformation Registration via Geodesics on a Learned Manifold of Images 680
Jihun Hamm, Christos Davatzikos, and Ragini Verma

A Non-rigid Registration Method for Serial microCT Mouse Hindlimb Images 688
Jung W. Suh, Dustin Scheinost, Donald P. Dione, Lawrence W. Dobrucki, Albert J. Sinusas, and Xenophon Papademetris

Non-rigid Image Registration with Uniform Gradient Spherical Patterns 696
Shu Liao and Albert C.S. Chung

Methods for Tractography-Driven Surface Registration of Brain Structures 705
Aleksandar Petrović, Stephen M. Smith, Ricarda A. Menke, and Mark Jenkinson

A Combined Surface And VOlumetric Registration (SAVOR) Framework to Study Cortical Biomarkers and Volumetric Imaging Data 713
Eli Gibson, Ali R. Khan, and Mirza Faisal Beg

Fast Tensor Image Morphing for Elastic Registration 721
Pew-Thian Yap, Guorong Wu, Hongtu Zhu, Weili Lin, and Dinggang Shen

DISCO: A Coherent Diffeomorphic Framework for Brain Registration under Exhaustive Sulcal Constraints 730
Guillaume Auzias, Joan Glaunès, Olivier Colliot, Matthieu Perrot, Jean-François Mangin, Alain Trouvé, and Sylvain Baillet

Evaluation of Lobar Biomechanics during Respiration Using Image Registration 739
Kai Ding, Youbing Yin, Kunlin Cao, Gary E. Christensen, Ching-Long Lin, Eric A. Hoffman, and Joseph M. Reinhardt

Evaluation of 4D-CT Lung Registration 747
Sven Kabus, Tobias Klinder, Keelin Murphy, Bram van Ginneken, Cristian Lorenz, and Josien P.W. Pluim

Slipping Objects in Image Registration: Improved Motion Field Estimation with Direction-Dependent Regularization 755
Alexander Schmidt-Richberg, Jan Ehrhardt, Rene Werner, and Heinz Handels

Multi-modal Registration Based Ultrasound Mosaicing 763
Oliver Kutter, Wolfgang Wein, and Nassir Navab

A Novel Method for Registration of US/MR of the Liver Based on the Analysis of US Dynamics 771
Sergiy Milko, Eivind Lyche Melvær, Eigil Samset, and Timor Kadir

Alignment of Viewing-Angle Dependent Ultrasound Images 779
Christian Wachinger and Nassir Navab

MR to Ultrasound Image Registration for Guiding Prostate Biopsy and Interventions 787
Yipeng Hu, Hashim Uddin Ahmed, Clare Allen, Doug Pendsé, Mahua Sahu, Mark Emberton, David Hawkes, and Dean Barratt

Combining Multiple True 3D Ultrasound Image Volumes through Re-registration and Rasterization 795
Songbai Ji, David W. Roberts, Alex Hartov, and Keith D. Paulsen

Biomechanically Constrained Groupwise US to CT Registration of the Lumbar Spine 803
Sean Gill, Parvin Mousavi, Gabor Fichtinger, Elvis Chen, Jonathan Boisvert, David Pichora, and Purang Abolmaesumi

A General PDE-Framework for Registration of Contrast Enhanced Images 811
Mehran Ebrahimi and Anne L. Martel

Statistically Deformable 2D/3D Registration for Accurate Determination of Post-operative Cup Orientation from Single Standard X-ray Radiograph 820
Guoyan Zheng

A Novel Intensity Similarity Metric with Soft Spatial Constraint for a Deformable Image Registration Problem in Radiation Therapy 828
Ali Khamene, Darko Zikic, Mamadou Diallo, Thomas Boettger, and Eike Rietzel

Intra-operative Multimodal Non-rigid Registration of the Liver for Navigated Tumor Ablation 837
Haytham Elhawary, Sota Oguro, Kemal Tuncali, Paul R. Morrison, Paul B. Shyn, Servet Tatli, Stuart G. Silverman, and Nobuhiko Hata

Neuroimage Analysis: Structure and Function

A Novel Measure of Fractional Anisotropy Based on the Tensor Distribution Function 845
Liang Zhan, Alex D. Leow, Siwei Zhu, Marina Barysheva, Arthur W. Toga, Katie L. McMahon, Greig I. De Zubicaray, Margaret J. Wright, and Paul M. Thompson

Iterative Co-linearity Filtering and Parameterization of Fiber Tracts in the Entire Cingulum 853
Marius de Groot, Meike W. Vernooij, Stefan Klein, Alexander Leemans, Renske de Boer, Aad van der Lugt, Monique M.B. Breteler, and Wiro J. Niessen

Think Global, Act Local; Projectome Estimation with BlueMatter 861
Anthony J. Sherbondy, Robert F. Dougherty, Rajagopal Ananthanarayanan, Dharmendra S. Modha, and Brian A. Wandell

Dual Tensor Atlas Generation Based on a Cohort of Coregistered non-HARDI Datasets 869
Matthan Caan, Caroline Sage, Maaike van der Graaf, Cornelis Grimbergen, Stefan Sunaert, Lucas van Vliet, and Frans Vos

Estimating Orientation Distribution Functions with Probability Density Constraints and Spatial Regularity 877
Alvina Goh, Christophe Lenglet, Paul M. Thompson, and René Vidal

Quantifying Brain Connectivity: A Comparative Tractography Study ... 886
Ting-Shuo Yo, Alfred Anwander, Maxime Descoteaux, Pierre Fillard, Cyril Poupon, and T.R. Knösche

Two-Tensor Tractography Using a Constrained Filter 894
James G. Malcolm, Martha E. Shenton, and Yogesh Rathi

Characterization of Anatomic Fiber Bundles for Diffusion Tensor Image Analysis 903
Rubén Cárdenes, Daniel Argibay-Quiñones, Emma Muñoz-Moreno, and Marcos Martin-Fernandez

A Riemannian Framework for Orientation Distribution Function Computing 911
Jian Cheng, Aurobrata Ghosh, Tianzi Jiang, and Rachid Deriche

Bias of Least Squares Approaches for Diffusion Tensor Estimation from Array Coils in DT–MRI 919
Antonio Tristán-Vega, Carl-Fredrik Westin, and Santiago Aja-Fernández

A Novel Global Tractography Algorithm Based on an Adaptive Spin Glass Model 927
Pierre Fillard, Cyril Poupon, and Jean-François Mangin

Tractography-Based Parcellation of the Cortex Using a Spatially-Informed Dimension Reduction of the Connectivity Matrix 935
Pauline Roca, Denis Rivière, Pamela Guevara, Cyril Poupon, and Jean-François Mangin

Belief Propagation Based Segmentation of White Matter Tracts in DTI 943
Pierre-Louis Bazin, John Bogovic, Daniel Reich, Jerry L. Prince, and Dzung L. Pham

Design and Construction of a Realistic DWI Phantom for Filtering Performance Assessment 951
Antonio Tristán-Vega and Santiago Aja-Fernández

Statistical Detection of Longitudinal Changes between Apparent Diffusion Coefficient Images: Application to Multiple Sclerosis 959
Hervé Boisgontier, Vincent Noblet, Félix Renard, Fabrice Heitz, Lucien Rumbach, and Jean-Paul Armspach

Tensor-Based Analysis of Genetic Influences on Brain Integrity Using DTI in 100 Twins 967
Agatha D. Lee, Natasha Leporé, Caroline Brun, Yi-Yu Chou, Marina Barysheva, Ming-Chang Chiang, Sarah K. Madsen, Greig I. de Zubicaray, Katie L. McMahon, Margaret J. Wright, Arthur W. Toga, and Paul M. Thompson

Robust Extrapolation Scheme for Fast Estimation of 3D Ising Field Partition Functions: Application to Within-Subject fMRI Data Analysis 975
Laurent Risser, Thomas Vincent, Philippe Ciuciu, and Jérôme Idier

Parcellation of fMRI Datasets with ICA and PLS-A Data Driven Approach 984
Yongnan Ji, Pierre-Yves Hervé, Uwe Aickelin, and Alain Pitiot

Adjusting the Neuroimaging Statistical Inferences for Nonstationarity 992
Gholamreza Salimi-Khorshidi, Stephen M. Smith, and Thomas E. Nichols

Using Real-Time fMRI to Control a Dynamical System by Brain Activity Classification 1000
Anders Eklund, Henrik Ohlsson, Mats Andersson, Joakim Rydell, Anders Ynnerman, and Hans Knutsson

Modeling Adaptation Effects in fMRI Analysis 1009
Wanmei Ou, Tommi Raij, Fa-Hsuan Lin, Polina Golland, and Matti Hämäläinen

A Cluster Overlap Measure for Comparison of Activations in fMRI Studies 1018
Guillermo A. Cecchi, Rahul Garg, and A. Ravishankar Rao

Author Index 1027

Table of Contents – Part II

Shape Modelling and Analysis

Building Shape Models from Lousy Data 1
Marcel Lüthi, Thomas Albrecht, and Thomas Vetter

Statistical Location Model for Abdominal Organ Localization 9
Jianhua Yao and Ronald M. Summers

Volumetric Shape Model for Oriented Tubular Structure from DTI Data 18
Hon Pong Ho, Xenophon Papademetris, Fei Wang, Hilary P. Blumberg, and Lawrence H. Staib

A Generic Probabilistic Active Shape Model for Organ Segmentation . . . 26
Andreas Wimmer, Grzegorz Soza, and Joachim Hornegger

Organ Segmentation with Level Sets Using Local Shape and Appearance Priors 34
Timo Kohlberger, M. Gökhan Uzunbaş, Christopher Alvino, Timor Kadir, Daniel O. Slosman, and Gareth Funka-Lea

Liver Segmentation Using Automatically Defined Patient Specific B-Spline Surface Models 43
Yi Song, Andy J. Bulpitt, and Ken W. Brodlie

Airway Tree Extraction with Locally Optimal Paths 51
Pechin Lo, Jon Sporring, Jesper Johannes Holst Pedersen, and Marleen de Bruijne

A Deformable Surface Model for Vascular Segmentation 59
Max W.K. Law and Albert C.S. Chung

A Deformation Tracking Approach to 4D Coronary Artery Tree Reconstruction 68
Yanghai Tsin, Klaus J. Kirchberg, Guenter Lauritsch, and Chenyang Xu

Automatic Extraction of Mandibular Nerve and Bone from Cone-Beam CT Data 76
Dagmar Kainmueller, Hans Lamecker, Heiko Seim, Max Zinser, and Stefan Zachow

Conditional Variability of Statistical Shape Models Based on Surrogate Variables 84
Rémi Blanc, Mauricio Reyes, Christof Seiler, and Gábor Székely

Surface/Volume-Based Articulated 3D Spine Inference through Markov Random Fields 92
Samuel Kadoury and Nikos Paragios

A Shape Relationship Descriptor for Radiation Therapy Planning 100
Michael Kazhdan, Patricio Simari, Todd McNutt, Binbin Wu, Robert Jacques, Ming Chuang, and Russell Taylor

Feature-Based Morphometry 109
Matthew Toews, William M. Wells III, D. Louis Collins, and Tal Arbel

Constructing a Dictionary of Human Brain Folding Patterns 117
Zhong Yi Sun, Matthieu Perrot, Alan Tucholka, Denis Rivière, and Jean-François Mangin

Topological Correction of Brain Surface Meshes Using Spherical Harmonics 125
Rachel Aine Yotter, Robert Dahnke, and Christian Gaser

Teichmüller Shape Space Theory and Its Application to Brain Morphometry 133
Yalin Wang, Wei Dai, Xianfeng Gu, Tony F. Chan, Shing-Tung Yau, Arthur W. Toga, and Paul M. Thompson

A Tract-Specific Framework for White Matter Morphometry Combining Macroscopic and Microscopic Tract Features 141
Hui Zhang, Suyash P. Awate, Sandhitsu R. Das, John H. Woo, Elias R. Melhem, James C. Gee, and Paul A. Yushkevich

Shape Modelling for Tract Selection 150
Jonathan D. Clayden, Martin D. King, and Chris A. Clark

Topological Characterization of Signal in Brain Images Using Min-Max Diagrams 158
Moo K. Chung, Vikas Singh, Peter T. Kim, Kim M. Dalton, and Richard J. Davidson

Particle Based Shape Regression of Open Surfaces with Applications to Developmental Neuroimaging 167
Manasi Datar, Joshua Cates, P. Thomas Fletcher, Sylvain Gouttard, Guido Gerig, and Ross Whitaker

Setting Priors and Enforcing Constraints on Matches for Nonlinear Registration of Meshes 175
Benoît Combès and Sylvain Prima

Parametric Representation of Cortical Surface Folding Based on Polynomials 184
Tuo Zhang, Lei Guo, Gang Li, Jingxin Nie, and Tianming Liu

Intrinsic Regression Models for Manifold-Valued Data 192
Xiaoyan Shi, Martin Styner, Jeffrey Lieberman, Joseph G. Ibrahim, Weili Lin, and Hongtu Zhu

Gender Differences in Cerebral Cortical Folding: Multivariate Complexity-Shape Analysis with Insights into Handling Brain-Volume Differences 200
Suyash P. Awate, Paul Yushkevich, Daniel Licht, and James C. Gee

Cortical Shape Analysis in the Laplace-Beltrami Feature Space 208
Yonggang Shi, Ivo Dinov, and Arthur W. Toga

Shape Analysis of Human Brain Interhemispheric Fissure Bending in MRI 216
Lu Zhao, Jarmo Hietala, and Jussi Tohka

Subject-Matched Templates for Spatial Normalization 224
Torsten Rohlfing, Edith V. Sullivan, and Adolf Pfefferbaum

Mapping Growth Patterns and Genetic Influences on Early Brain Development in Twins 232
Yasheng Chen, Hongtu Zhu, Dinggang Shen, Hongyu An, John Gilmore, and Weili Lin

Tensor-Based Morphometry with Mappings Parameterized by Stationary Velocity Fields in Alzheimer's Disease Neuroimaging Initiative 240
Matías Nicolás Bossa, Ernesto Zacur, and Salvador Olmos

Motion Analysis, Physical Based Modelling and Image Reconstruction

High-Quality Model Generation for Finite Element Simulation of Tissue Deformation 248
Orcun Goksel and Septimiu E. Salcudean

Biomechanically-Constrained 4D Estimation of Myocardial Motion 257
Hari Sundar, Christos Davatzikos, and George Biros

Predictive Simulation of Bidirectional Glenn Shunt Using a Hybrid Blood Vessel Model 266
Hao Li, Wee Kheng Leow, and Ing-Sh Chiu

Surgical Planning and Patient-Specific Biomechanical Simulation for Tracheal Endoprostheses Interventions 275
Miguel A. González Ballester, Amaya Pérez del Palomar, José Luís López Villalobos, Laura Lara Rodríguez, Olfa Trabelsi, Frederic Pérez, Ángel Ginel Cañamaque, Emilia Barrot Cortés, Francisco Rodríguez Panadero, Manuel Doblaré Castellano, and Javier Herrero Jover

Mesh Generation from 3D Multi-material Images 283
Dobrina Boltcheva, Mariette Yvinec, and Jean-Daniel Boissonnat

Interactive Simulation of Flexible Needle Insertions Based on Constraint Models 291
Christian Duriez, Christophe Guébert, Maud Marchal, Stephane Cotin, and Laurent Grisoni

Real-Time Prediction of Brain Shift Using Nonlinear Finite Element Algorithms 300
Grand Roman Joldes, Adam Wittek, Mathieu Couton, Simon K. Warfield, and Karol Miller

Model-Based Estimation of Ventricular Deformation in the Cat Brain 308
Fenghong Liu, S. Scott Lollis, Songbai Ji, Keith D. Paulsen, Alexander Hartov, and David W. Roberts

Contact Studies between Total Knee Replacement Components Developed Using Explicit Finite Elements Analysis 316
Lucian Gheorghe Gruionu, Gabriel Gruionu, Stefan Pastrama, Nicolae Iliescu, and Taina Avramescu

A Hybrid 1D and 3D Approach to Hemodynamics Modelling for a Patient-Specific Cerebral Vasculature and Aneurysm 323
Harvey Ho, Gregory Sands, Holger Schmid, Kumar Mithraratne, Gordon Mallinson, and Peter Hunter

Incompressible Cardiac Motion Estimation of the Left Ventricle Using Tagged MR Images 331
Xiaofeng Liu, Khaled Z. Abd-Elmoniem, and Jerry L. Prince

Vibro-Elastography for Visualization of the Prostate Region: Method Evaluation 339
Seyedeh Sara Mahdavi, Mehdi Moradi, Xu Wen, William J. Morris, and Septimiu E. Salcudean

Modeling Respiratory Motion for Cancer Radiation Therapy Based on Patient-Specific 4DCT Data 348
Jaesung Eom, Chengyu Shi, Xie George Xu, and Suvranu De

Correlating Chest Surface Motion to Motion of the Liver Using ε-SVR – A Porcine Study 356
Floris Ernst, Volker Martens, Stefan Schlichting, Armin Beširević, Markus Kleemann, Christoph Koch, Dirk Petersen, and Achim Schweikard

Respiratory Motion Estimation from Cone-Beam Projections Using a Prior Model 365
Jef Vandemeulebroucke, Jan Kybic, Patrick Clarysse, and David Sarrut

Heart Motion Abnormality Detection via an Information Measure and Bayesian Filtering 373
Kumaradevan Punithakumar, Shuo Li, Ismail Ben Ayed, Ian Ross, Ali Islam, and Jaron Chong

Automatic Image-Based Cardiac and Respiratory Cycle Synchronization and Gating of Image Sequences 381
Hari Sundar, Ali Khamene, Liron Yatziv, and Chenyang Xu

Dynamic Cone Beam Reconstruction Using a New Level Set Formulation 389
Andreas Keil, Jakob Vogel, Günter Lauritsch, and Nassir Navab

Spatio-temporal Reconstruction of dPET Data Using Complex Wavelet Regularisation 398
Andrew McLennan and Michael Brady

Evaluation of q-Space Sampling Strategies for the Diffusion Magnetic Resonance Imaging 406
Haz-Edine Assemlal, David Tschumperlé, and Luc Brun

On the Blurring of the Funk–Radon Transform in Q–Ball Imaging 415
Antonio Tristán-Vega, Santiago Aja-Fernández, and Carl-Fredrik Westin

Multiple Q-Shell ODF Reconstruction in Q-Ball Imaging 423
Iman Aganj, Christophe Lenglet, Guillermo Sapiro, Essa Yacoub, Kamil Ugurbil, and Noam Harel

Neuro, Cell and Multiscale Image Analysis

Lossless Online Ensemble Learning (LOEL) and Its Application to Subcortical Segmentation 432
Jonathan H. Morra, Zhuowen Tu, Arthur W. Toga, and Paul M. Thompson

Improved Maximum a Posteriori Cortical Segmentation by Iterative Relaxation of Priors 441
Manuel Jorge Cardoso, Matthew J. Clarkson, Gerard R. Ridgway, Marc Modat, Nick C. Fox, and Sebastien Ourselin

Anatomically Informed Bayesian Model Selection for fMRI Group Data Analysis 450
Merlin Keller, Marc Lavielle, Matthieu Perrot, and Alexis Roche

A Computational Model of Cerebral Cortex Folding 458
Jingxin Nie, Gang Li, Lei Guo, and Tianming Liu

Tensor-Based Morphometry of Fibrous Structures with Application to Human Brain White Matter 466
Hui Zhang, Paul A. Yushkevich, Daniel Rueckert, and James C. Gee

A Fuzzy Region-Based Hidden Markov Model for Partial-Volume Classification in Brain MRI 474
Albert Huang, Rafeef Abugharbieh, and Roger Tam

Brain Connectivity Using Geodesics in HARDI 482
Mickaël Péchaud, Maxime Descoteaux, and Renaud Keriven

Functional Segmentation of fMRI Data Using Adaptive Non-negative Sparse PCA (ANSPCA) 490
Bernard Ng, Rafeef Abugharbieh, and Martin J. McKeown

Genetics of Anisotropy Asymmetry: Registration and Sample Size Effects 498
Neda Jahanshad, Agatha D. Lee, Natasha Leporé, Yi-Yu Chou, Caroline C. Brun, Marina Barysheva, Arthur W. Toga, Katie L. McMahon, Greig I. de Zubicaray, Margaret J. Wright, and Paul M. Thompson

Extending Genetic Linkage Analysis to Diffusion Tensor Images to Map Single Gene Effects on Brain Fiber Architecture 506
Ming-Chang Chiang, Christina Avedissian, Marina Barysheva, Arthur W. Toga, Katie L. McMahon, Greig I. de Zubicaray, Margaret J. Wright, and Paul M. Thompson

Vascular Territory Image Analysis Using Vessel Encoded Arterial Spin Labeling 514
Michael A. Chappell, Thomas W. Okell, Peter Jezzard, and Mark W. Woolrich

Predicting MGMT Methylation Status of Glioblastomas from MRI Texture 522
Ilya Levner, Sylvia Drabycz, Gloria Roldan, Paula De Robles, J. Gregory Cairncross, and Ross Mitchell

Tumor Invasion Margin on the Riemannian Space of Brain Fibers 531
Dana Cobzas, Parisa Mosayebi, Albert Murtha, and Martin Jagersand

A Conditional Random Field Approach for Coupling Local Registration with Robust Tissue and Structure Segmentation 540
Benoit Scherrer, Florence Forbes, and Michel Dojat

Robust Atlas-Based Brain Segmentation Using Multi-structure Confidence-Weighted Registration 549
Ali R. Khan, Moo K. Chung, and Mirza Faisal Beg

Discriminative, Semantic Segmentation of Brain Tissue in MR Images 558
Zhao Yi, Antonio Criminisi, Jamie Shotton, and Andrew Blake

Use of Simulated Atrophy for Performance Analysis of Brain Atrophy Estimation Approaches 566
Swati Sharma, Vincent Noblet, François Rousseau, Fabrice Heitz, Lucien Rumbach, and Jean-Paul Armspach

Fast and Robust 3-D MRI Brain Structure Segmentation 575
Michael Wels, Yefeng Zheng, Gustavo Carneiro, Martin Huber, Joachim Hornegger, and Dorin Comaniciu

Multiple Sclerosis Lesion Segmentation Using an Automatic Multimodal Graph Cuts 584
Daniel García-Lorenzo, Jeremy Lecoeur, Douglas L. Arnold, D. Louis Collins, and Christian Barillot

Towards Accurate, Automatic Segmentation of the Hippocampus and Amygdala from MRI 592
D. Louis Collins and Jens C. Pruessner

An Object-Based Method for Rician Noise Estimation in MR Images . . . 601
Pierrick Coupé, José V. Manjón, Elias Gedamu, Douglas Arnold, Montserrat Robles, and D. Louis Collins

Cell Segmentation Using Front Vector Flow Guided Active Contours . . . 609
Fuhai Li, Xiaobo Zhou, Hong Zhao, and Stephen T.C. Wong

Segmentation and Classification of Cell Cycle Phases in Fluorescence Imaging 617
Ilker Ersoy, Filiz Bunyak, Vadim Chagin, M. Christina Cardoso, and Kannappan Palaniappan

Steerable Features for Statistical 3D Dendrite Detection 625
Germán González, François Aguet, François Fleuret, Michael Unser, and Pascal Fua

Graph-Based Pancreatic Islet Segmentation for Early Type 2 Diabetes Mellitus on Histopathological Tissue 633
Xenofon Floros, Thomas J. Fuchs, Markus P. Rechsteiner, Giatgen Spinas, Holger Moch, and Joachim M. Buhmann

Detection of Spatially Correlated Objects in 3*D* Images Using Appearance Models and Coupled Active Contours 641
Kishore Mosaliganti, Arnaud Gelas, Alexandre Gouaillard, Ramil Noche, Nikolaus Obholzer, and Sean Megason

Intra-retinal Layer Segmentation in Optical Coherence Tomography Using an Active Contour Approach 649
Azadeh Yazdanpanah, Ghassan Hamarneh, Benjamin Smith, and Marinko Sarunic

Mapping Tissue Optical Attenuation to Identify Cancer Using Optical Coherence Tomography 657
Robert A. McLaughlin, Loretta Scolaro, Peter Robbins, Christobel Saunders, Steven L. Jacques, and David D. Sampson

Analysis of MR Images of Mice in Preclinical Treatment Monitoring of Polycystic Kidney Disease 665
Stathis Hadjidemetriou, Wilfried Reichardt, Martin Buechert, Juergen Hennig, and Dominik von Elverfeldt

Actin Filament Tracking Based on Particle Filters and Stretching Open Active Contour Models 673
Hongsheng Li, Tian Shen, Dimitrios Vavylonis, and Xiaolei Huang

Image Analysis and Computer Aided Diagnosis

Toward Early Diagnosis of Lung Cancer 682
Ayman El-Baz, Georgy Gimel'farb, Robert Falk, Mohamed Abou El-Ghar, Sabrina Rainey, David Heredia, and Teresa Shaffer

Lung Extraction, Lobe Segmentation and Hierarchical Region Assessment for Quantitative Analysis on High Resolution Computed Tomography Images 690
James C. Ross, Raúl San José Estépar, Alejandro Díaz, Carl-Fredrik Westin, Ron Kikinis, Edwin K. Silverman, and George R. Washko

Learning COPD Sensitive Filters in Pulmonary CT 699
Lauge Sørensen, Pechin Lo, Haseem Ashraf, Jon Sporring, Mads Nielsen, and Marleen de Bruijne

Automated Anatomical Labeling of Bronchial Branches Extracted from CT Datasets Based on Machine Learning and Combination Optimization and Its Application to Bronchoscope Guidance 707
Kensaku Mori, Shunsuke Ota, Daisuke Deguchi, Takayuki Kitasaka, Yasuhito Suenaga, Shingo Iwano, Yosihnori Hasegawa, Hirotsugu Takabatake, Masaki Mori, and Hiroshi Natori

Multi-level Ground Glass Nodule Detection and Segmentation in CT Lung Images 715
Yimo Tao, Le Lu, Maneesh Dewan, Albert Y. Chen, Jason Corso, Jianhua Xuan, Marcos Salganicoff, and Arun Krishnan

Global and Local Multi-valued Dissimilarity-Based Classification: Application to Computer-Aided Detection of Tuberculosis 724
Yulia Arzhaeva, Laurens Hogeweg, Pim A. de Jong, Max A. Viergever, and Bram van Ginneken

Noninvasive Imaging of Electrophysiological Substrates in Post Myocardial Infarction 732
Linwei Wang, Heye Zhang, Ken C.L. Wong, Huafeng Liu, and Pengcheng Shi

Unsupervised Inline Analysis of Cardiac Perfusion MRI 741
Hui Xue, Sven Zuehlsdorff, Peter Kellman, Andrew Arai, Sonia Nielles-Vallespin, Christophe Chefdhotel, Christine H. Lorenz, and Jens Guehring

Pattern Recognition of Abnormal Left Ventricle Wall Motion in Cardiac MR 750
Yingli Lu, Perry Radau, Kim Connelly, Alexander Dick, and Graham Wright

Septal Flash Assessment on CRT Candidates Based on Statistical Atlases of Motion 759
Nicolas Duchateau, Mathieu De Craene, Etel Silva, Marta Sitges, Bart H. Bijnens, and Alejandro F. Frangi

Personalized Modeling and Assessment of the Aortic-Mitral Coupling from 4D TEE and CT 767
Razvan Ioan Ionasec, Ingmar Voigt, Bogdan Georgescu, Yang Wang, Helene Houle, Joachim Hornegger, Nassir Navab, and Dorin Comaniciu

A New 3-D Automated Computational Method to Evaluate In-Stent Neointimal Hyperplasia in In-Vivo Intravascular Optical Coherence Tomography Pullbacks 776
Serhan Gurmeric, Gozde Gul Isguder, Stéphane Carlier, and Gozde Unal

MKL for Robust Multi-modality AD Classification 786
Chris Hinrichs, Vikas Singh, Guofan Xu, and Sterling Johnson

A New Approach for Creating Customizable Cytoarchitectonic Probabilistic Maps without a Template 795
Amir M. Tahmasebi, Purang Abolmaesumi, Xiujuan Geng, Patricia Morosan, Katrin Amunts, Gary E. Christensen, and Ingrid S. Johnsrude

A Computer-Aided Diagnosis System of Nuclear Cataract via Ranking 803
Wei Huang, Huiqi Li, Kap Luk Chan, Joo Hwee Lim, Jiang Liu, and Tien Yin Wong

Automated Segmentation of the Femur and Pelvis from 3D CT Data of Diseased Hip Using Hierarchical Statistical Shape Model of Joint Structure 811
Futoshi Yokota, Toshiyuki Okada, Masaki Takao, Nobuhiko Sugano, Yukio Tada, and Yoshinobu Sato

Computer-Aided Assessment of Anomalies in the Scoliotic Spine in 3-D MRI Images 819
Florian Jäger, Joachim Hornegger, Siegfried Schwab, and Rolf Janka

Optimal Graph Search Segmentation Using Arc-Weighted Graph for Simultaneous Surface Detection of Bladder and Prostate 827
Qi Song, Xiaodong Wu, Yunlong Liu, Mark Smith, John Buatti, and Milan Sonka

Automated Calibration for Computerized Analysis of Prostate Lesions Using Pharmacokinetic Magnetic Resonance Images 836
Pieter C. Vos, Thomas Hambrock, Jelle O. Barenstz, and Henkjan J. Huisman

Spectral Embedding Based Probabilistic Boosting Tree (ScEPTre): Classifying High Dimensional Heterogeneous Biomedical Data 844
Pallavi Tiwari, Mark Rosen, Galen Reed, John Kurhanewicz, and Anant Madabhushi

Automatic Correction of Intensity Nonuniformity from Sparseness of Gradient Distribution in Medical Images 852
Yuanjie Zheng, Murray Grossman, Suyash P. Awate, and James C. Gee

Weakly Supervised Group-Wise Model Learning Based on Discrete Optimization 860
René Donner, Horst Wildenauer, Horst Bischof, and Georg Langs

Image Segmentation and Analysis

ECOC Random Fields for Lumen Segmentation in Radial Artery IVUS Sequences 869
Francesco Ciompi, Oriol Pujol, Eduard Fernández-Nofrerías, Josepa Mauri, and Petia Radeva

Dynamic Layer Separation for Coronary DSA and Enhancement in Fluoroscopic Sequences 877
Ying Zhu, Simone Prummer, Peng Wang, Terrence Chen, Dorin Comaniciu, and Martin Ostermeier

An Inverse Scattering Algorithm for the Segmentation of the Luminal Border on Intravascular Ultrasound Data 885
E. Gerardo Mendizabal-Ruiz, George Biros, and Ioannis A. Kakadiaris

Image-Driven Cardiac Left Ventricle Segmentation for the Evaluation of Multiview Fused Real-Time 3-Dimensional Echocardiography Images 893
Kashif Rajpoot, J. Alison Noble, Vicente Grau, Cezary Szmigielski, and Harald Becher

Left Ventricle Segmentation via Graph Cut Distribution Matching 901
Ismail Ben Ayed, Kumaradevan Punithakumar, Shuo Li, Ali Islam, and Jaron Chong

Combining Registration and Minimum Surfaces for the Segmentation of the Left Ventricle in Cardiac Cine MR Images 910
Marie-Pierre Jolly, Hui Xue, Leo Grady, and Jens Guehring

Left Ventricle Segmentation Using Diffusion Wavelets and Boosting 919
Salma Essafi, Georg Langs, and Nikos Paragios

3D Cardiac Segmentation Using Temporal Correlation of Radio Frequency Ultrasound Data 927
Maartje M. Nillesen, Richard G.P. Lopata, Henkjan J. Huisman, Johan M. Thijssen, Livia Kapusta, and Chris L. de Korte

Improved Modelling of Ultrasound Contrast Agent Diminution for Blood Perfusion Analysis 935
Christian Kier, Karsten Meyer-Wiethe, Günter Seidel, and Alfred Mertins

A Novel 3D Joint Markov-Gibbs Model for Extracting Blood Vessels from PC–MRA Images 943
Ayman El-Baz, Georgy Gimel'farb, Robert Falk, Mohamed Abou El-Ghar, Vedant Kumar, and David Heredia

Atlas-Based Improved Prediction of Magnetic Field Inhomogeneity for Distortion Correction of EPI Data 951
Clare Poynton, Mark Jenkinson, and William Wells III

3D Prostate Segmentation in Ultrasound Images Based on Tapered and Deformed Ellipsoids 960
Seyedeh Sara Mahdavi, William J. Morris, Ingrid Spadinger, Nick Chng, Orcun Goksel, and Septimiu E. Salcudean

An Interactive Geometric Technique for Upper and Lower Teeth Segmentation 968
Binh Huy Le, Zhigang Deng, James Xia, Yu-Bing Chang, and Xiaobo Zhou

Enforcing Monotonic Temporal Evolution in Dry Eye Images 976
Tamir Yedidya, Peter Carr, Richard Hartley, and Jean-Pierre Guillon

Ultrafast Localization of the Optic Disc Using Dimensionality Reduction of the Search Space 985
Ahmed Essam Mahfouz and Ahmed S. Fahmy

Using Frankenstein's Creature Paradigm to Build a Patient Specific Atlas 993
Olivier Commowick, Simon K. Warfield, and Grégoire Malandain

Atlas-Based Automated Segmentation of Spleen and Liver Using Adaptive Enhancement Estimation 1001
Marius George Linguraru, Jesse K. Sandberg, Zhixi Li, John A. Pura, and Ronald M. Summers

A Two-Level Approach towards Semantic Colon Segmentation: Removing Extra-Colonic Findings 1009
Le Lu, Matthias Wolf, Jianming Liang, Murat Dundar, Jinbo Bi, and Marcos Salganicoff

Segmentation of Lumbar Vertebrae Using Part-Based Graphs and Active Appearance Models 1017
Martin G. Roberts, Tim F. Cootes, Elisa Pacheco, Teik Oh, and Judith E. Adams

Utero-Fetal Unit and Pregnant Woman Modeling Using a Computer Graphics Approach for Dosimetry Studies 1025
Jérémie Anquez, Tamy Boubekeur, Lazar Bibin, Elsa Angelini, and Isabelle Bloch

Cross Modality Deformable Segmentation Using Hierarchical Clustering and Learning 1033
Yiqiang Zhan, Maneesh Dewan, and Xiang Sean Zhou

3D Multi-branch Tubular Surface and Centerline Extraction with 4D Iterative Key Points 1042
Hua Li, Anthony Yezzi, and Laurent Cohen

Multimodal Prior Appearance Models Based on Regional Clustering of Intensity Profiles 1051
François Chung and Hervé Delingette

3D Medical Image Segmentation by Multiple-Surface Active Volume Models 1059
Tian Shen and Xiaolei Huang

Model Completion via Deformation Cloning Based on an Explicit Global Deformation Model 1067
Qiong Han, Stephen E. Strup, Melody C. Carswell, Duncan Clarke, and Williams B. Seales

Supervised Nonparametric Image Parcellation 1075
Mert R. Sabuncu, B.T. Thomas Yeo, Koen Van Leemput, Bruce Fischl, and Polina Golland

Thermal Vision for Sleep Apnea Monitoring 1084
Jin Fei, Ioannis Pavlidis, and Jayasimha Murthy

Tissue Tracking in Thermo-physiological Imagery through Spatio-temporal Smoothing 1092
Yan Zhou, Panagiotis Tsiamyrtzis, and Ioannis T. Pavlidis

Depth Data Improves Skin Lesion Segmentation 1100
Xiang Li, Ben Aldridge, Lucia Ballerini, Robert Fisher, and Jonathan Rees

A Fully Automatic Random Walker Segmentation for Skin Lesions in a Supervised Setting 1108
Paul Wighton, Maryam Sadeghi, Tim K. Lee, and M. Stella Atkins

Author Index 1117

Optimal Transseptal Puncture Location for Robot-Assisted Left Atrial Catheter Ablation

Jagadeesan Jayender[1], Rajni V. Patel[2], Gregory F. Michaud[3], and Nobuhiko Hata[1]

[1] Department of Radiology, Harvard Medical School, Brigham and Women's Hospital, USA
jayender@bwh.harvard.edu, hata@bwh.harvard.edu
[2] Department of Electrical and Computer Engineering and Department of Surgery The University of Western Ontario, Canadian Surgical Technologies and Advanced Robotics (CSTAR), Canada
rajni.patel@lhsc.on.ca
[3] Department of Cardiology, Harvard Medical School, Brigham and Women's Hospital, USA
gfmichaud@partners.org.

Abstract. The preferred method of treatment for Atrial Fibrillation (AF) is by catheter ablation wherein a catheter is guided into the left atrium through a transseptal puncture. However, the transseptal puncture constrains the catheter, thereby limiting its maneuverability and increasing the difficulty in reaching various locations in the left atrium. In this paper, we address the problem of choosing the optimal transseptal puncture location for performing cardiac ablation to obtain maximum maneuverability of the catheter. We have employed an optimization algorithm to maximize the Global Isotropy Index (GII) to evaluate the optimal transseptal puncture location. As part of this algorithm, a novel kinematic model for the catheter has been developed based on a continuum robot model. Preoperative MR/CT images of the heart are segmented using the open source image-guided therapy software, Slicer 3, to obtain models of the left atrium and septal wall. These models are input to the optimization algorithm to evaluate the optimal transseptal puncture location. Simulation results for the optimization algorithm are presented in this paper.

1 Introduction

Atrial fibrillation (AF) is a type of arrhythmia where the atria beat irregularly and out of coordination with the ventricles. When rate and rhythm control drugs are ineffective in patients suffering from AF, catheter ablation is performed as a standard method of treatment. More recently, a robotics based master-slave approach for performing cardiac ablation has been introduced. At our institution, we have direct clinical experience with the Sensei robot from Hansen Medical Inc. [1]. Although the robot is capable of precise catheter motions and maintaining stable contact with the heart wall, the clinician often encounters

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 1–8, 2009.

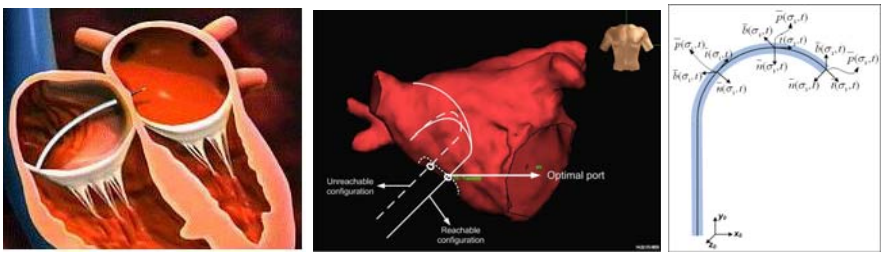


Fig. 1. (a) Catheter entering the LA through the transseptal puncture, (b) Difficulty in maneuvering the catheter due to improper transseptal puncture location, (c) Model for the catheter

difficulty in maneuvering the catheter and approaching the desired point on the heart at the required angle. The lack of dexterity of the catheter arises from the constraint imposed on the motion of the catheter due to the transseptal puncture [2], see Figure 1(a). The interatrial septal wall effectively acts like a Remote Center of Motion (RCM) about which the catheter is manipulated. This can severely affect the dexterity and maneuverability of the catheter (see Figure 1(b)), thereby affecting the clinician's performance. In addition, there is also danger of perforating the aortic arch with catastrophic results due to accidental needle puncture [3]. Therefore, it is of prime importance for the clinician to puncture the transseptal wall at an optimal point to maximize the dexterity of the catheter in the left atrium while minimizing the possibility of puncturing the aortic arch. In this paper, we develop a computational algorithm to determine the optimal transseptal puncture location for performing cardiac ablation. Once determined, the puncture location will be registered to the Intracardiac Echo (ICE) image so that the clinician can guide the needle in real-time to the optimal puncture location to obtain maximum maneuverability of the catheter, while avoiding puncturing of the aorta. The problem of choosing the location of the transseptal puncture bears similarity to the problem of choosing the ports on the thoracic cavity for insertion of laparoscopic tools and endoscopes to perform Robot-Assisted Minimally Invasive Surgery (RA-MIS). Several researchers have studied the problem of choosing the optimal ports to maximize the performance of the robot (and surgeon) to perform RA-MIS [4]-[7]. In this paper we have adopted the Global Isotropy Index (GII) [8], [5] to define the performance measure for quantifying the dexterity of the catheter within the left atrium. We have developed an optimization algorithm to maximize the GII for efficiently maneuvering the catheter in the left atrium.

2 Modeling of Catheter

The catheter is considered as being made up of infinitesimal rigid links along a backbone curve. The backbone curve is defined in terms of the Frenet-Serret frame. The curve is represented in the parametric form, $\bar{x} = \bar{x}(s,t)$, where s is the

parameter which represents the curve length and t is the time. The Frenet-Serret frame is defined at each point $\sigma(s,t)$ along the backbone curve and consists of the tangent $\bar{t}(s,t)$, normal $\bar{n}(s,t)$ and binormal $\bar{b}(s,t)$ vector at point $\sigma(s,t)$, as shown in Figure 1(c). At any point σ along the curve $\bar{x}(s,t)$, the local frame can be defined as $\Phi(\sigma,t)$. In terms of the Euler angles, the frame ${}^0\Phi(\sigma,t)$ can be defined with respect to the base frame ${}^0\Phi(0,t)$ as a rotation of α about $\bar{b}(\sigma)$ and β about $\bar{n}(\sigma)$ vector. Also the angles α and β can be defined in terms of the curvature and torsion as follows

$$\alpha = \int_0^s \varsigma(\sigma,t)d\sigma \qquad \beta = \int_0^s \tau(\sigma,t)d\sigma \tag{1}$$

In this case, we assume that the curvature ς and torsion τ are constant. Therefore, ${}^0\Phi(\sigma,t)$ can be written as

$${}^0\Phi(\sigma,t) = \mathtt{Rot}(\bar{b},\alpha)\mathtt{Rot}(\bar{n},\beta)\ i.e., \tag{2}$$

$${}^0\Phi(\sigma) = \begin{bmatrix} \cos(\sigma\varsigma)\cos(\sigma\tau) & -\sin(\sigma\varsigma) & \cos(\sigma\varsigma)\sin\sigma\tau \\ \sin(\sigma\varsigma)\cos(\sigma\tau) & \cos(\sigma\varsigma) & \sin(\sigma\varsigma)\sin(\sigma\tau) \\ -\sin(\sigma\tau) & 0 & \cos(\sigma\tau) \end{bmatrix} \tag{3}$$

The position vector $\bar{p}(\sigma,t)$ of a point σ on the curve relative to the origin $\bar{p}(0)$ can be computed by integrating infinitesimal curve lengths along the tangent vector. In order words, ${}^0\bar{p}(\sigma)$, which represents the position of a point σ on the curve as viewed in the base frame ${}^0\Phi(0)$ is given by

$${}^0\bar{p}(\sigma,t) = \int_0^s {}^0\Phi(\eta,t)\hat{e_x}d\eta \tag{4}$$

Each segment of the catheter within the heart can be considered to consist of two rotational joints and a prismatic joint. The joint angle vector can be written as $\bar{\theta} = [0\ \varsigma\ \tau]^t$ and the translational vector can be written as $\bar{d} = [l_x\ 0\ 0]^t$ The rotational velocity for a joint in the local Frenet-Serret frame with respect to the base frame can be written as

$${}^0\omega(\sigma) = \int_0^\sigma \dot{\bar{\theta}}^\sigma \Phi(\nu)d\nu \mathtt{(for\ a\ rotational\ joint)} \tag{5}$$

$${}^0\omega(\sigma) = 0 \mathtt{(for\ a\ prismatic\ joint)} \tag{6}$$

Similarly the translational velocities for the rotational and prismatic joints can be written as

$${}^0v(\sigma) = \int_0^\sigma \dot{\bar{\theta}}^\sigma \Phi(\nu) \times ({}^\sigma p(\nu,t) - {}^\sigma p(\nu,t))d\nu \mathtt{(for\ a\ rotational\ joint)} \tag{7}$$

$${}^0v(\sigma) = \int_0^\sigma \dot{\bar{d}}^\sigma \Phi(\nu)d\nu \mathtt{(for\ a\ prismatic\ joint)} \tag{8}$$

Equations (5)-(8) together can be written in a compact form as

$$\begin{bmatrix} {}^0v(\sigma) \\ {}^0\omega(\sigma) \end{bmatrix} = \int_0^\sigma \begin{bmatrix} {}^\sigma\Phi(\nu,t) & [p(\nu,t) - p(\sigma,t)\times]\Phi(\nu,t) \\ 0 & \Phi(\nu,t) \end{bmatrix} \bar{A} \begin{bmatrix} \dot{d} \\ \dot{\varsigma} \\ \dot{\tau} \end{bmatrix} d\nu \tag{9}$$

where

$$\bar{A} = \begin{bmatrix} 1\,0\,0\,0\,0\,0 \\ 0\,0\,0\,0\,0\,1 \\ 0\,0\,0\,0\,1\,0 \end{bmatrix}^T \quad \texttt{and}[a\times] \triangleq \begin{bmatrix} 0 & -a_z & a_y \\ a_z & 0 & -a_x \\ -a_y & a_x & 0 \end{bmatrix} \tag{10}$$

Using the standard robotics terminology, the Jacobian operator can be defined as

$$\mathcal{J}(\sigma,t) = \int_0^\sigma \begin{bmatrix} {}^\sigma\Phi(\nu,t) & [p(\nu,t) - p(\sigma,t)\times]\Phi(\nu,t) \\ 0 & \Phi(\nu,t) \end{bmatrix} \bar{A}(.)d\nu \tag{11}$$

From (4), we can estimate the position of the catheter with respect to the base coordinates for a given configuration of the catheter. However, for the optimization algorithm, we will assume that the catheter is in contact with a certain point inside the left atrium while being constrained at the transseptal puncture location. Therefore, this gives us the base and end-effector coordinates of the catheter. It is required to solve the inverse kinematics problem to evaluate the configuration of the catheter for the distal end of the catheter to be in contact with a particular point inside the heart. The problem of solving the inverse kinematics of the catheter is formulated as a dynamical problem which requires only the computation of the forward kinematics, as determined by (4). Let us represent the solution of the inverse kinematics problem as $\hat{q}(t)$ corresponding to a trajectory $\hat{x}(t)$ which satisfies the forward kinematics given by (4). Let $e(t)$ represent the error between the desired Cartesian position $\hat{x}(t)$ and the actual Cartesian position obtained from the state variable q of the iteration algorithm. The error dynamics can be written as

$$\dot{e}(t) = \dot{\hat{x}}(t) - \dot{x}(t) = \dot{\hat{x}}(t) - \mathcal{J}\dot{q} \tag{12}$$

We choose a purely proportional control law to solve for $\dot{q}$ as given by

$$\dot{q} = \alpha \mathcal{J}^T e \tag{13}$$

It has been shown in [9] that by choosing a control law as given by (13), the error e is bounded and can be made small with an appropriate choice of α, with the added benefit of less computational complexity.

3 Global Isotropy Index

The Jacobian matrix relates the end-effector frame velocities $\dot{x}$ and forces f to the corresponding joint rates $\dot{q}$ and torques τ, as given by the following equations

$$\dot{x} = \mathcal{J}\dot{q} \tag{14}$$

$$\tau = \mathcal{J}^T f \tag{15}$$

Condition number κ of the Jacobian $\mathcal{J}$ can be considered as the error amplification factor from the joint space to the Cartesian space. Taking the norm on both sides of (14), we obtain

$$\frac{\|\delta x\|}{\|x\|} \leq \|\mathcal{J}^{-1}\|\|\mathcal{J}\|\frac{\|\delta q\|}{\|q\|} \tag{16}$$

The condition number κ is, therefore, defined as [10]

$$\kappa(J) = \|\mathcal{J}^{-1}\|\|\mathcal{J}\| = \frac{\overline{\sigma}(J(q))}{\underline{\sigma}(J(q))} \tag{17}$$

where $\overline{\sigma}$ and $\underline{\sigma}$ represent the maximum and minimum singular values of $\mathcal{J}$. The condition number represents the dexterity of the robot at a specified point and often, the inverse of the condition number is used since the value of $\kappa(J)^{-1} \in [0, 1]$. For a robot in a singular configuration, $\underline{\sigma} = 0$, implying that for $\kappa(J)^{-1} = 0$. The closer $\kappa(J)^{-1}$ or $\kappa(J)$ is to unity, the more accurate and dextrous the robot is. A configuration of the robot is defined as an isotropic configuration, which is the most dextrous configuration for the robot, when the condition number $\kappa(J)$ corresponding to this configuration is unity. Therefore, in order to ensure that the robot (in our case the catheter) is well controlled in all regions of the left atrium, we try to maximize the dexterity or a measure of isotropy. However, it should be noted that the Jacobian is a function of the local position (in turn joint variables), therefore the condition number is a local measure of the dexterity of the robot at any position. The transseptal puncture location should be chosen such that the dexterity of the catheter is maximized at all points in the left atria. Therefore, there is a need to define the performance measure as a global measure rather than a local measure. In [8], the Global Isotropy Index (GII) was proposed to define the overall dexterity of the robot throughout the workspace. The GII is defined as

$$GII(y_i) = \frac{\min\limits_{x_0 \in W_i} \underline{\sigma}(\mathcal{J}(x_0))}{\max\limits_{x_1 \in W_i} \overline{\sigma}(\mathcal{J}(x_1))} \tag{18}$$

where $W_i = W(y_i) = \{x_k - y_i : x_k \in W_a\}$, $y_i \in W_s$. For our application, $W_a \in \Re^3$ is the set of points on the left atrium and $W_s \in \Re^3$ is the set of points on the septal wall. It should be noted that GII is a global measure as compared to the condition number which is a local measure. In addition, it should also be noted that the GII is related to the inverse of the condition number rather than the condition number itself, in order to restrict the values to the range $[0, 1]$. A GII of 1 implies that the catheter is isotropic at every point in the workspace and behaves uniformly in all directions, which is the ideal condition.

Since linear and circumferential lesions may be needed at most points in the left atrium depending on the type of atrial fibrillation (persistant, paroxysmal or permanent), we consider the surgical workspace to be the entire left atrium. The workspace can be further reduced based on the clinician's input in the preoperative stage. The problem is to choose the transseptal puncture location to maximize the dexterity of the catheter while accessing the surgical workspace. The dexterity of the catheter in the surgical workspace is quantified in terms of the GII. The optimization algorithm searches for the point $y_{opt} \in W_s$, which maximizes the GII, corresponding to the optimal location for the transseptal puncture. This location constrains the catheter and can be considered as the base coordinates of the catheter. The distal end of the catheter (considered as the end-effector) touches the points on the left atrium. For each point on the left atrium, the configuration of the catheter is estimated using inverse kinematics

and the corresponding Jacobian $\mathcal{J}$ and singular values $(\overline{\sigma}, \underline{\sigma})$ are evaluated and provided to the optimization algorithm.

4 Experimental Results

A pre-operative CT image was first loaded in Slicer 3. The CT image is a contrast enhanced image consisting of 64 slices. The resolution of each slice is 0.7mm×0.7mm×0.7mm. An expert radiologist was given the task of manually segmenting the left atrium, the septal wall and the right atrium. The models were then generated in Slicer 3 and the corresponding VTK file consisting of the data points of the model was created. The file consisting of the coordinates of the left atrium and the septal wall were input to the optimization algorithm. The optimal transseptal puncture location algorithm was implemented using MATLAB on an Intel Core2 Duo 2.00GHz machine with 1GB RAM. Figure 2 (a) and (b) show the results of the variation of GII as a function of x and y coordinates respectively. The minimum value of GII was 0.0112 and the maximum value was 0.4362, a location which provides the best isotropy and in turn dexterity of the catheter. In addition, the GII values were converted to a scalar VIBGYOR color map (red representing low GII and violet representing high GII) corresponding to the points in the VTK file and displayed on the model of the septal wall in Slicer, as shown in Figure 3(a). Since the catheter is constrained to the transseptal puncture point, the catheter loses some dexterity and this can be noticed in the low values of the GII at most points on the septal wall, represented by the red color in Figure 3(a). The mean value of GII on the septal wall was 0.0327. Figure 3(b) shows the location of the optimal puncture with respect to the left and right atria. Figure 4(a) shows the mesh plot of the GII on the septal wall. The colorbar reflects the values of the GII at different locations on the septal wall.

In Figure 4(b), we have compared the condition number κ for six points on the left atria corresponding to two transseptal puncture locations S_1 and S_2. It can be seen that although the GII corresponding to the point S_1 is larger than that

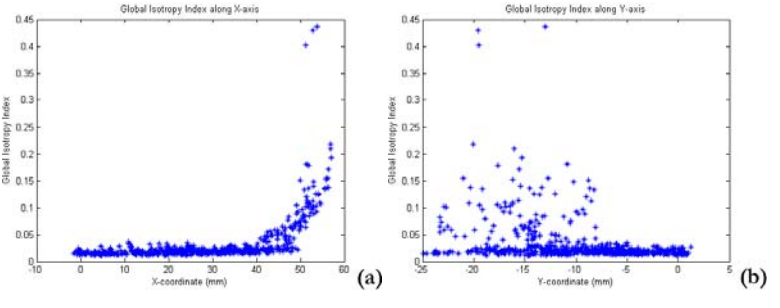


Fig. 2. (a) Graph showing the variation of GII as a function of the x-coordinate, (b) Graph showing the variation of GII as a function of the y-coordinate

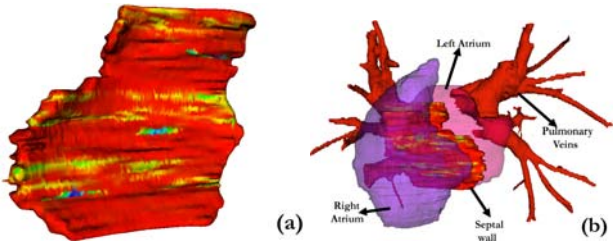


Fig. 3. (a) Magnified view of the septal wall with the VIBGYOR color map representing the GII at various locations. Red represents points with low GII while blue represents points with high GII; (b) Segmented model of the LA, RA, PV and septal wall with the color map representing the GII at different locations.

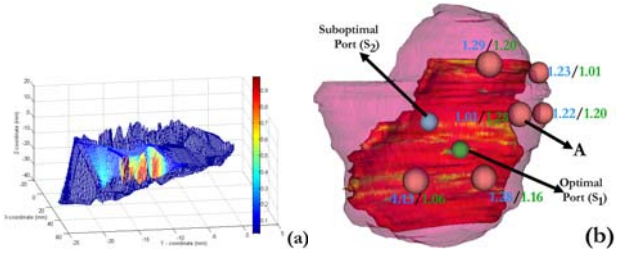


Fig. 4. (a) Mesh plot showing the color map overlaid over the septal wall with colorbar representing the GII, (b) 3D view of 2 points chosen on the septal wall and 6 points on LA. The condition number corresponding to each point is represented as x_1/x_2, where x_1 corresponds to the condition number of the point assuming that the catheter is passing through the suboptimal port S_2 and x_2 corresponds to the condition number of the point assuming that the catheter is passing through the optimal port S_1.

corresponding to S_2, the condition number at point A is smaller (1.01) for S_2 compared to the condition number of 1.28 for S_1. This implies that the control of the catheter is more accurate at point A, if S_2 is chosen as the transseptal puncture location. As mentioned earlier, GII is a global measure of the dexterity or isotropy of the catheter at all points in the left atria while the condition number is a local measure. It should also be noted that the result of the transseptal puncture algorithm is highly specific to the patient and is dependent on the shape and size of the left atrium, position and number of pulmonary veins and the shape and size of the septal wall.

5 Conclusion

The objective of this work was to determine the optimal transseptal puncture location and assist the clinician in guiding the needle in real-time to this location. In this paper, we have developed an optimization algorithm based on the Global

Isotropy Index (GII) to evaluate the optimal position of the transseptal puncture for left atrial cardiac ablation. As part of this algorithm, we have extended the continuum robot model to describe the kinematics of the catheter within the left atrium. Based on this model, the configuration of the catheter was evaluated such that the catheter makes contact with the desired points on the left atrium with constraints imposed on its motion by the transseptal puncture. The choice of the transseptal puncture location affects the uniformity of catheter manipulation. The optimal puncture location ensures maximum dexterity of the catheter within the left atrium and also ensures that the catheter has the capability of reaching various locations in the heart. Successful implementation of this algorithm in clinical practice will eventually lead to reducing the time necessary to complete the procedure, improving access to difficult regions in the left atrium, reducing the amount of manipulations of the catheter required to reach a point in the left atrium and minimizing the amount of fatigue for the clinicians.

References

1. Hansen Medical Inc., http://www.hansenmedical.com/
2. Pappone, C., Vicedomini, G., Manguso, F., Gugliotta, F., Mazzone, P., Gulletta, S., Sora, N., Sala, S., Marzi, A., Augello, G.: Robotic magnetic navigation for atrial fibrillation ablation. Journal of the American College of Cardiology 47, 1390–1400 (2006)
3. Ponti, R.D., et al.: Trans-septal catheterization in the electrophysiology laboratory: data from a multicenter survey spanning 12 years. J. Am. Coll. Cardiol. 47, 1037–1042 (2006)
4. Cannon, J., Stoll, J., Selha, S., Dupont, P., Howe, R., Torchiana, D.: Port placement planning in robot-assisted coronary artery bypass. IEEE Transactions on Robotics and Automation 19, 912–917 (2003)
5. Trejos, A.L., Patel, R.V., Ross, I., Kiaii, B.: Optimizing port placement for robot-assisted minimally invasive cardiac surgery. Int. J. Med. Robotics Comput. Assist. Surg. 3, 355–364 (2007)
6. Adhami, L., Coste-Manière, È., Boissonnat, J.-D.: Planning and simulation of robotically assisted minimal invasive surgery. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 624–633. Springer, Heidelberg (2000)
7. Badani, K.K., Muhletaler, F., Fumo, M., Kaul, S., Peabody, J.O., Bhandari, M., Menon, M.: Optimizing robotic renal surgery: The lateral camera port placement technique and current results. Journal of Endourology 22, 507–510 (2008)
8. Stocco, L., Salcudean, S.E., Sassani, F.: Fast constrained global minimax optimization of robot parameters. Robotica 16, 595–605 (1998)
9. Sciavicco, L., Siciliano, B.: A solution algorithm to the inverse kinematic problem for redundant manipulators. IEEE Transactions on Robotics and Automation 4, 403–410 (1988)
10. Merlet, J.P.: Jacobian, manipulability, condition number, and accuracy of parallel robots. Journal of Mechanical Design 128, 199–206 (2006)

Towards Guidance of Electrophysiological Procedures with Real-Time 3D Intracardiac Echocardiography Fusion to C-arm CT

Wolfgang Wein[1], Estelle Camus[2], Matthias John[3], Mamadou Diallo[1], Christophe Duong[2], Amin Al-Ahmad[4], Rebecca Fahrig[4], Ali Khamene[1], and Chenyang Xu[1]

[1] Imaging & Visualization Dept., Siemens Corporate Research, Princeton, NJ, USA
[2] Siemens Healthcare, Ultrasound Business Unit, Mountain View, CA, USA
[3] Siemens Healthcare, AX Business Unit, Forchheim, Germany
[4] Department of Radiology, Stanford University, CA, USA

Abstract. This paper describes a novel method for improving the navigation and guidance of devices and catheters in electrophysiology and interventional cardiology procedures using volumetric data fusion. The clinical workflow includes the acquisition and reconstruction of CT data from a C-arm X-ray angiographic system and the real-time acquisition of volumetric ultrasound datasets with a new intracardiac real-time 3D ultrasound catheter. Mono- and multi-modal volumetric registration methods, as well as visualization modes, that are suitable for real-time fusion are described, which are the key components of this work. Evaluation on phantom and in-vivo animal data shows that it is feasible to register and track the motion of real-time 3D intracardiac ultrasound in C-arm CT.

1 Introduction

Pre- and intra-procedural imaging is an integral part of many cardiac interventions, both surgical and minimally invasive (percutaneous) ones. Regarding high-resolution static pre- and intra-operative imaging, Computed Tomography (CT), including C-arm CT, is most widely used for diagnosis and planning. However these imaging modalities do not have real-time capabilities. As a real-time interventional modality, intra-cardiac echocardiography (ICE) is now widely used in many cathlabs, e.g. for guidance of left atrial ablation procedures.

We have developed an intra-cardiac ultrasound catheter (Siemens ACUSON AcuNav^{TM} V), which is capable of acquiring ultrasound volumes from within the heart in real-time [1]. This novel imaging modality may, on its own, have an tremendous impact on intra-cardiac procedures. However, even 2D ICE technology, when used in the electrophysiology (EP) environment, has been found to require adequate training and experience in order to fully exploit its benefits. Finding the orientation of the ICE catheter within the body and recognizing anatomical structures can be challenging, especially when the catheter is moved quickly. The 3D capability of our new system allows fusion of real-time ICE

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 9–16, 2009.

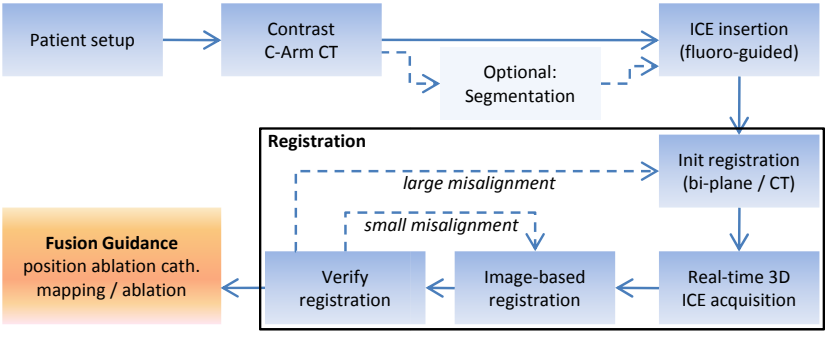


Fig. 1. Workflow overview

data with pre-operative or intra-operative high-resolution CT or MRI images, potentially reducing procedure time and improving treatment outcome. Baseline anatomical images can be combined with ICE imaging, which provides real-time monitoring of morphological and functional changes of cardiac tissue.

The CartoSound module (Biosense Webster, Israel) allows to superimpose 2D real-time ICE with an integrated magnetic position sensor to pre-operative information. Related work regarding registration and overall fusion setup is described in [2] (using heart surface points as registration features), and [3]. In the latter, an ECG- and respiration-gated ICE sweep is registered to CT, by manual initialization and image-based refinement. Thereafter, the position sensor update allows for live fusion of the 2D ICE with superimposed CT slices and segmented structures.

In our work, we would like to take advantage of the capabilities of the novel real-time 3D ICE catheter. In order to minimize set-up time, increase flexibility, and ensure compatibility to any EP mapping solution used in clinical practice we target a purely image-based technique for registration.

2 Methods

2.1 Proposed Clinical Workflow

In the following, we assume the case of an ablation procedure to treat atrial fibrillation. The proposed workflow (as outlined in figure 1) can also be used e.g. for guidance of transseptal punctures or guidance of intracorporeal devices during cardiac interventions.

At the beginning of the procedure, a baseline cardiac C-arm CT reconstruction with contrast agent injection is performed. Optionally, automatic or semi-automatic segmentation of the heart chambers (in particular of the left atrium) and the esophagus may be performed. Then, the ICE catheter is inserted into the body and guided to the heart (typically to the right side via the femoral or jugular vein), if needed, under fluoroscopic guidance. The ICE-CT registration

is initialized by locating the ICE catheter tip either in a bi-plane fluoro view, or in a low-dose C-arm CT reconstruction. This data is in turn automatically registered to the baseline volume. Next, an ICE acquisition is performed, typically by rotating the ICE catheter such that it covers the whole left atrium. This recording is automatically registered to the baseline CT, using the algorithms described below. After visual assessment, this process can be repeated depending on the quality of the alignment. Thereafter, the actual EP procedure, which includes electromapping and ablation can be executed under fusion imaging. The position of the ICE information within the C-arm CT volume can be updated any time by image-based tracking and/or additional fluoroscopic acquisitions.

In the next section, we describe algorithms supporting this workflow by providing image-based registration and tracking of real-time 3D ICE within C-arm CT.

2.2 Cardiac Phase Detection

Since the baseline CT data is only available in one cardiac phase (e.g. end-diastole), only volumes of the live ICE data corresponding to that phase can be considered for registration. ECG gating is a common solution; however the ECG signal might be delayed with respect to actual cardiac motion, or, in the worst case, not directly be related at all, in patients with heart arrhythmia. An image-based technique for detecting periodic motion has therefore been chosen that uses extrema of the cumulated phase shift in the spectral domain of successive image frames [4]. End-diastole ICE frames are tagged with this method and then used for image-based registration to the baseline data[1].

2.3 Registration of Single Volume

In order to allow image-based registration of ICE to C-arm CT, the modalities have to be made more similar in a pre-processing step, and then a nonlinear optimization refines the transformation parameters with respect to a measure of the alignment quality. For CT-ultrasound registration in the abdomen, the state-of-the-art uses a simulation of ultrasonic echogenicity and reflection from CT, and a specific similarity metric to overcome their unknown combination [5]. In the case of echocardiography, only the reflection term is significant, showing the myocardial walls of the cardiac cavities. The blood inside the chambers results in completely hypoechoic regions. In addition, the diffusivity of the reflections at cardiac walls is relatively high, and shadowing does not usually occur within our anatomy of interest (i.e. imaging the left atrium from the right atrium). We therefore drop the simulated echogenicity and orientation-dependency from the approach in [5], and precompute a global simulated ultrasound reflection volume, henceforth denoted R, from the baseline C-arm CT volume. The registration of a single end-diastole ICE volume U to the baseline volume is then described as

$$\arg\max_{\phi} \mathrm{CC}\left(U, R \circ T^{-1}(\phi)\right) \tag{1}$$

[1] We would like to thank Hari Sundar from SCR for providing the phase component.

where T is the rigid homogeneous transformation matrix mapping U on R, parameterized by a 6-vector ϕ. CC is the Pearson's correlation coefficient, which we use because it is independent of scale and bias between the simulated reflection and actual ultrasound intensities. $R \circ T^{-1}$ denotes a resampling of R with the transformation T^{-1}. The scan lines of the ICE data are scan converted into a cartesian grid, which both serves as a low-pass filter and makes the resolution more isotropic.

An initialization is provided by manually extracting the position of the ICE catheter from a bi-plane fluoroscopic acquisition, or a low-dose C-arm CT reconstruction. The optimization then consists of a global search on the rotation around the catheter axis, followed by a Simplex direct search method.

2.4 Trajectory Registration

A single ICE volume provides a limited field of view within the cardiac anatomy, potentially resulting in multiple ambiguous locations suggesting good alignment. Because the physician is maneuvering the ICE catheter within the heart to image different anatomical regions of interest, successive end-diastole ICE volumes can be used to create an extended field-of-view volume. These volumes need to be brought into correct alignment themselves. We simultaneously optimize mono- and multi-modal similarity metrics for that purpose. Now the pose parameters of a sequence of successive ICE frames are refined, until both the alignment to the baseline C-arm CT, as well as their internal alignment are optimal:

$$\arg\max_{\phi} \frac{1}{N}\sum_{i=1}^{N} \mathrm{CC}(U_i, R \circ T_i^{-1}) - \frac{\alpha}{N-1}\sum_{i=1}^{N-1} \mathrm{SAD}(U_i, U_{i+1} \circ (T_i^{-1} T_{i+1})) \quad (2)$$

As before, the correlation coefficient serves as multimodal measure. Sum of Absolute Differences $\mathrm{SAD}(U,V) = \frac{1}{|\Omega|}\sum |u_i - v_i|$ works best for measuring the alignment of successive ICE volumes, where the flexibility of CC is not needed. Since it is a dissimilarity measure, we negate and weight it with a parameter α. Rather than optimizing the unconstrained problem with $N * 6$ degrees of freedom (DOF), we manipulate only the first and last transformation, assuming all volumes are located on a linear trajectory:

$$\phi = \begin{pmatrix} \phi_0 \\ \phi_1 \end{pmatrix}; \quad T_i = T(\phi_0) \exp\left(\frac{i-1}{N-1} \log(T(\phi_0)^{-1} T(\phi_1))\right) \quad (3)$$

Here, Lie manifold based interpolation of transformation matrices is used; the optimization problem is reduced to 12 DOF. Note that we do not have to weight the individual SAD values with respect to volume overlap, because by using this manifold interpolation all successive volumes will have equal overlap.

For arbitrarily long ICE sequences, or continuous motion tracking, the registration described by equation 3, is repeatedly applied to an overlapping subset of all volumes.

2.5 Implementation Details

We use a software framework developed in C++, OpenGL and GLSL. As is common nowadays, we have accelerated scan-conversion, volume resampling, similarity measure computation, and visualization, by implementing them on a graphics processing unit (GPU). The incoming ICE volumes are immediately uploaded to a ringbuffer of 3D textures in graphics memory.

The registration of a single catheter motion with 10 ICE volumes takes $\approx 0.6s$ (where each ICE volume is used with $\approx$ 130K voxels) on an Nvidia GeForce 8800 GTX GPU, compared to $27s$ in a reference software implementation.

For enhancing the visual perception of the ICE data, an efficient shading technique is used for volume rendering, that does not require 3D volume gradients [6]. The real-time visualization comprises three blended multi-planar reformatted (MPR) slices of the C-arm CT volume and ICE superimposed in color, cropped 3D volume rendering, as well as mesh visualization of an optional four-chamber segmentation of the heart, see figure 4(b).

3 Experiments

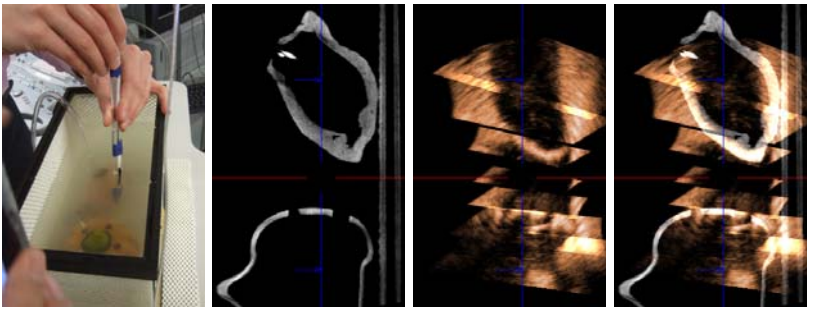
3.1 Phantom Data

We conducted phantom experiments with the Siemens ACUSON AcuNav V ICE catheter submerged in an acrylic water container, which was cushioned with rubber material to minimize acoustic reflections at the walls. It was placed on the patient table of a Siemens Axiom Artis angiographic suite, such that we could acquire 3D C-arm CT (software: *syngo*© DynaCT Cardiac) at any time. We attached a magnetic tracking sensor (Ascension MicroBird) to the ICE catheter and recorded both static and dynamic sequences on clay phantoms (a left atrium model and a small Halloween pumpkin, the latter chosen because of its sharp facial imaging features, but also seasonally inspired).

In order to use the magnetic tracking information, spatial calibration has to be performed to derive the location of the actual ICE catheter image with respect to the tracking sensor's coordinate system. Because of the unique imaging characteristics of the novel real-time 3D ICE catheter none of the off-the-shelf calibration techniques seemed appropriate. We decided to use a simultaneous optimization of the calibration pose ϕ_C and registration ϕ_R, comparing 12 ICE volumes from a wide sweep across the clay objects against the C-arm CT volume:

$$\arg\max_{\phi} \sum_{i=1}^{N} \mathrm{CC}\left(U_i, R \circ (T_R(\phi_1) T_T^i T_C(\phi_0))^{-1}\right); \quad \phi = \begin{pmatrix} \phi_C \\ \phi_R \end{pmatrix} \tag{4}$$

where $T_C(\phi_C)$ is the calibration transformation (ICE to sensor), T_T^i the tracking matrix (sensor to tracker reference), and $T_R(\phi_R)$ the registration transformation (tracker reference to CT). R is here replaced with the original C-arm CT volume, because for such water/clay data the CT measurements correspond directly to ultrasonic reflection. We verifed the calibration with a follow-up C-arm CT of



(a) Setup (b) CT (c) ICE (d) Overlay

Fig. 2. Evaluation on phantom data

the ICE catheter mounted in a still configuration. The result allowed us to depict both static and dynamic real-time 3D ICE sequences fused with the respective C-arm CT data (figure 2). We then evaluated the trajectory registration method using the tracked and calibrated ICE as Ground Truth, on 18 volumes from another sweep over both clay objects. For many algorithm executions, using $7-12$ ICE volumes at a time for minimizing equation 3, advancing $1-4$ volumes in between, resulted in a mean accuracy in translation and rotation of $2.2mm$ and $1.8°$ respectively over the entire motion. The weighting α interestingly has minor influence, as long as it is not chosen to be extremely large. In that case, the multi-modal component is ignored, and the registration of successive frames creates a drift (error at last frame $7.4mm$, $7.4°$).

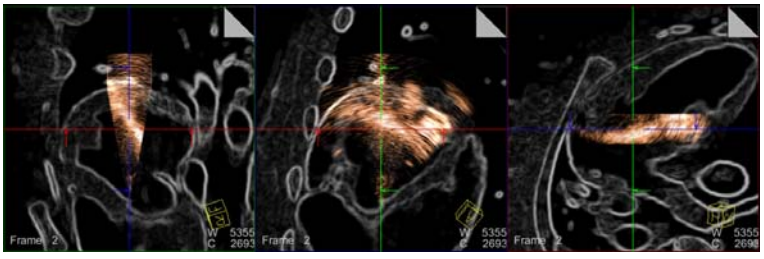
3.2 Animal Studies

We are currently evaluating the real-time 3D ICE imaging modality itself and our approach for ICE-CT fusion (works in progress) in a series of animal and human studies. In the following, we report on initial results from data acquired during an acute pig study[2].

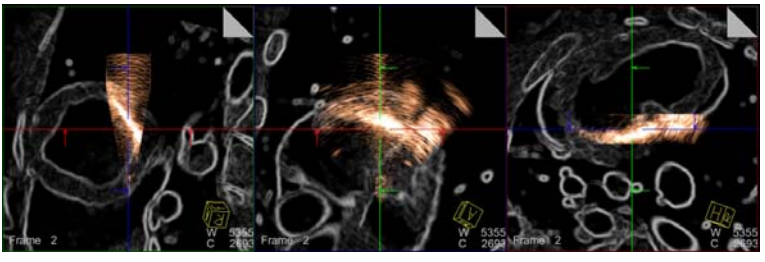
The baseline C-arm CT reconstruction was executed in forced breath-hold and cardiac gating. All ICE recordings were done during normal respiration. The used ICE catheter did not have a position sensor integrated in its housing. Therefore, we defined quasi-ground truth registration on the recorded ICE sequences, using all the advanced visualization at our disposal (figure 4). Figure 3 depicts a comparison of physician-defined ground truth and the result of our image-based registration (section 2.3). For four data sets (showing aortic valve, left atrium, left ventricle), the average discrepancy between expert-defined and image-based registration was $16mm$ and $11°$.

We also evaluated the trajectory registration on a recording where the ICE catheter was rotated within the right atrium. The first volume was roughly aligned by a physician, and the rotation direction of the trajectory indicated by

[2] The procedure was performed on a 35-kg pig according to a protocol approved by Stanford University's Institutional Animal Care and Use Committee.



(a) Ground Truth



(b) after automatic registration

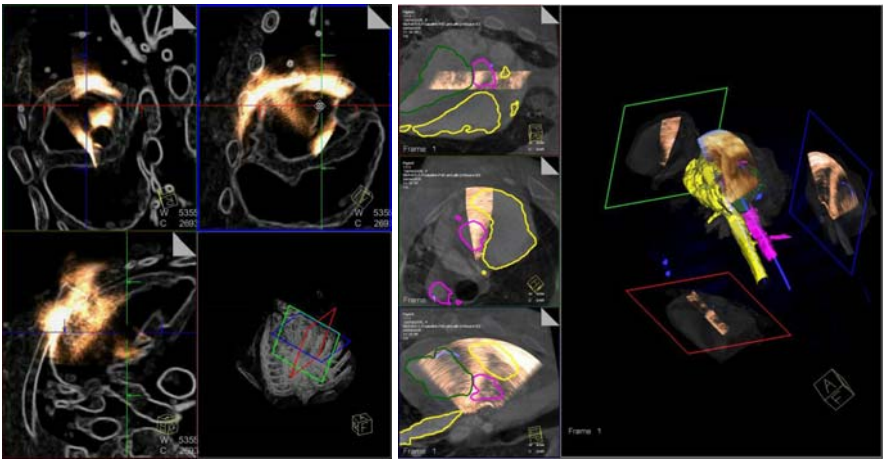
Fig. 3. Cross-sections of CT reflection overlayed with single end-diastole ICE volume

manually dragging the series of end-diastole volumes. The automatic registration then nicely pulls the whole sequence into good visual alignment, see figure 4(a).

4 Conclusion

We presented the first image fusion results with a novel real-time 3D intra-cardiac ultrasound catheter. We showed that automatic registration to C-arm CT data, and appropriate fused visualization for guiding EP procedures, are possible, after a rough registration initialization. The underlying technological novelties are the use of an image-based detection of cardiac phase, an algorithm for simultaneous mono- and multi-modal registration of a trajectory of ICE volumes, and dedicated real-time visualization. Furthermore, we detailed a workflow, which could enable such a fusion system to be used in clinical practice in the future, with all necessary imaging integrated in the interventional suite.

We evaluated our methods quantitatively on phantom data, and qualitatively on in-vivo pig data. The trajectory registration technique can be used to track cardiac end-diastole ultrasound volumes within a C-arm CT volume in real-time, without additional position tracking hardware. A particular advantage is that our registration is directly based on the anatomical structures in the two modalities, which inherently compensates for residual cardiac and respiratory motion. Given the limited field of view offered by the ICE catheter, accurate registration of 3D ICE volumes with CT datasets is a difficult task. Therefore, an extended-field-of-view acquisition, as proposed, is crucial for non-ambiguous, more accurate alignment.



(a) Trajectory Registration (b) 3D Visualization

Fig. 4. Left: Trajectory registration result of a rotation in a pig's left atrium, with ICE (yellow) superimposed onto the CT gradient volume in 3 MPRs, and their corresponding location in 3D. Right: 3D visualization of an ICE volume with C-arm CT and outlines from a four-chamber segmentation (with 3 MPRs in the left column).

Further experimental investigations are needed in order to create proper ground truth registration and refine our method in terms of robust, contiguous tracking anywhere in the heart. Particularly, the limitations of the concept of a linear motion trajectory needs to be assessed more precisely in the case of the ICE catheter, and could be addressed with more powerful motion models.

References

1. Wilser, W., Barnes, S., Garbini, L.: Helical acoustic array for medical ultrasound. United States Patent No. 20,080,125,659 (May 2008)
2. Zhong, H., Kanade, T., Schwartzman, D.: Virtual touch: An efficient registration method for catheter navigation in left atrium. In: MICCAI 2006 proceedings (2006)
3. Sun, Y., Kadoury, S., Li, Y., John, M., Resnick, J., Plambeck, G., Liao, R., Sauer, F., Xu, C.: Image guidance of intracardiac ultrasound with fusion of pre-operative images. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 60–67. Springer, Heidelberg (2007)
4. Sundar, H., Khamene, A., Yatziv, L., Wein, W., Xu, C.: Automatic image-based cardiac and respiratory cycle synchronization and gating of image sequences. In: Yang, G.-Z., et al. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 381–388. Springer, Heidelberg (2009)
5. Wein, W., Brunke, S., Khamene, A., Callstrom, M., Navab, N.: Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis 12, 577–585 (2008)
6. Sumanaweera, T.: Applying real-time shading to 3D ultrasound visualization. GPU Gems, ch. 40 (2004)

Personalized Pulmonary Trunk Modeling for Intervention Planning and Valve Assessment Estimated from CT Data

Dime Vitanovski[2,3], Razvan Ioan Ionasec[1,5,⋆], Bogdan Georgescu[1], Martin Huber[2], Andrew Mayall Taylor[4], Joachim Hornegger[3], and Dorin Comaniciu[1]

[1] Integrated Data Systems, Siemens Corporate Research, Princeton, USA
[2] Software and Engineering, Siemens Corporate Technology, Erlangen, Germany
[3] Chair of Pattern Recognition, Friedrich-Alexander-University, Erlangen, Germany
[4] UCL Institute of Child Health & Great Ormond Street Hospital for Children
[5] Computer Aided Medical Procedures, Technical University Munich, Germany
razvan.ionasec@siemens.com

Abstract. Pulmonary valve disease affects a significant portion of the global population and often occurs in conjunction with other heart dysfunctions. Emerging interventional methods enable percutaneous pulmonary valve implantation, which constitute an alternative to open heart surgery. As minimal invasive procedures become common practice, imaging and non-invasive assessment techniques turn into key clinical tools. In this paper, we propose a novel approach for intervention planning as well as morphological and functional quantification of the pulmonary trunk and valve. An abstraction of the anatomic structures is represented through a four-dimensional, physiological model able to capture large pathological variation. A hierarchical estimation, based on robust learning methods, is applied to identify the patient-specific model parameters from volumetric CT scans. The algorithm involves detection of piecewise affine parameters, fast centre-line computation and local surface delineation. The estimated personalized model enables for efficient and precise quantification of function and morphology. This ability may have impact on the assessment and surgical interventions of the pulmonary valve and trunk. Experiments performed on 50 cardiac computer tomography sequences demonstrated the average speed of 202 seconds and accuracy of 2.2mm for the proposed approach. An initial clinical validation yielded a significant correlation between model-based and expert measurements. To the best of our knowledge this is the first dynamic model of the pulmonary trunk and right ventricle outflow track estimated from CT data.

1 Introduction

Valvular heart disease (VHD) is an important cardiac disorder that affects a large number of patients and often requires operative intervention. In most cases, pulmonary abnormality occurs in conjunction with other heart diseases and can

⋆ Corresponding author.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 17–25, 2009.

be caused by congenital defects and pulmonary hypertension endocarditis. Such conditions require constant monitoring and a complex clinical workflow which includes: patient evaluation, percutaneous intervention planning, valve replacement and repair and follow-up studies.

Until recently, the pulmonary valve replacement has been performed surgically on open heart [1], with all associated risks: high morbidity, incidence of neurological damage, stroke and repeated valve replacement [2]. Minimal invasive procedures for the pulmonary valve are less traumatic, reducing valve replacement risk [3]. Percutaneous pulmonary valve implantation (PPVI) [4] is a new developed technique for transcatheter placement of a valve stent. The main difficulties of PPVI are: the assessment of the pulmonary trunk and the right ventricle outflow track (RVOT $<$ 22mm) before treatment [4], the classification of patients suitable for the procedure [5] and identification of the exact location for anchoring the stent [6]. Hence, precise assessment of the morphology and dynamics is crucial for the pre-procedural planning and successful intervention of PPVI.

Cardiac CT imaging is often performed when high spatial resolution, soft tissue contrast or dynamics is essential. The key advantage is the ability to acquire multiple non-invasive and accurate scans required for evaluation. In standard clinical settings these are the imaging techniques used to gain the information about the shape of the RVOT and the pulmonary artery. The acquired data (4D CT) is usually translated into sets of 2D planes for manual quantification and visual evaluation due to the lack of appropriate methods and tools for processing 3D/4D information. Measurements are tedious to obtain and moreover known to be affected by inaccuracies, as 2D alignment and sectioning is ambiguous and might lead to misinterpretation and distensibility [4].

A dynamic personalized model is expected to enhance quantification accuracy and non-invasive visualization of the human pulmonary trunk and RVOT. Recently, Schievano et. al. [5] proposed a patient specific 3D model manual extracted from MRI data. Although it may positively impact patient evaluation and procedure planning, the modeling requires a long time, the accuracy is influenced by user experience, and does not include dynamics.

Within this paper we present a dynamic model of the pulmonary trunk automatically estimated from CT data. The segmentation task is performed by fitting a 4D generic model to the patient specific image data. The proposed workflow, from modeling to quantification, involves three stages. In the first place, a generic model of the pulmonary artery is computed from a comprehensive training set (see Section 2). This is fitted to the available volumetric scans using robust machine-learning methods in order to obtain a patient specific physiological model (Section 3). Finally, the morphology and function of the artery is efficiently quantified from the personalized abstract representation (Section 4.1).

Extensive experiments are performed on a data set of 50 CT patients (503 Volumes), which includes a diverse range of morphological and pathological variation and patients' age. Results of the experiments demonstrated the average

accuracy of 2.2mm and speed of 202 seconds for the proposed approach on cardiac CT data (Section 4.2). An initial clinical validation performed demonstrates a significant correlation between model-based and manually performed expert measurements.

2 Model Representation

In order to accurately represent morphology and dynamics, our model design is consistent with the anatomy and physiology of the pulmonary trunk. The architecture of the model is anatomically oriented and includes all clinical relevant structures. Geometries are represented parametrically using non-uniform rational B-Splines (NURBS)[7], which is standard mathematical formulation in computation modeling.

Initially, the pulmonary trunk is represented by five anatomical landmarks: Trigone (L_t), RVOT (L_{rvot}), Main-Bifurcation (L_{mb}), Left-Bifurcation (L_{lb}) and Right-Bifurcation (L_{rb}). The Trigone defines the splitting point between the right ventricle outflow track and the tricuspid valve plane. The RVOT landmark is located on the same plane as the Trigone, parallel to the tricuspid valve plane. Main-Bifurcation defines the center of the trunk branching, while Left-Bifurcation and Right-Bifurcation mark the branches' end points (see Fig. 1(a)).

Constrained by the landmarks, the artery's center lines form the second abstraction layer (see Fig. 1(b)). The Main Center Line is bounded by the RVOT and Bifurcation. The Left and the Right Center Lines extend from the main bifurcation landmark to the Left-Bifurcation and Right-Bifurcation landmark, respectively.

The main pulmonary artery along with its bifurcations is modeled as parametric closed surface similar to a deformed cylinder. NURBS are used as the parametric representation due to the compact mathematical formulation and convenient properties. The topology and morphology of the arterial structure is characterized and constrained by the previously defined landmarks and centerlines. In order to obtain the physiological model of the pulmonary artery, a cylindrical structure along with the center line has been computed (see Fig. 1(c)).

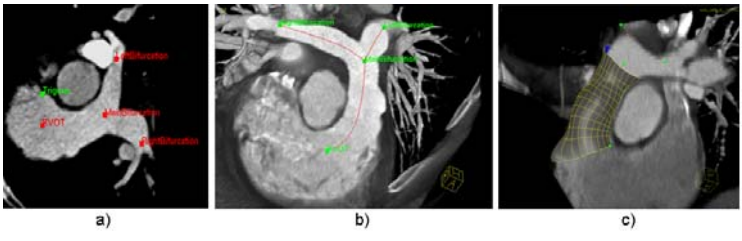


Fig. 1. Model representation

3 Model Estimation

The abstract parameters of the pulmonary trunk proposed in Section 2 are determined for a specific patient from the available data (four dimensional CT). A specific instance of the model is constrained in four-dimensional Euclidean space by 5 landmarks, 3 centerlines (determined by 27 discrete points) and 3 surfaces (determined by 966 control points) cumulating into 5*T[1] + 3*T*(9 + 322) parameters. The high dimensionality of the parameter space makes the direct estimation in the original space a very difficult task. Therefore, we proposed a four-step approach by first detecting a piecewise similarity transformation, followed by landmark, center line and full dynamic model detection.

3.1 Similarity Transformation and Landmark Detection

For each landmark except the triagone, which is included into the RVOT box θ_{rvot}), we define a corresponding bounding box, parameterized by an affine transformation $\theta = (X, Y, Z, \alpha, \beta, \chi, S_x, S_y, S_z)$. Each box characterizes one anatomical structure: RVOT plane is defined by θ_{rvot} with orientation given by the center line's tangent, θ_{mb} defines the main bifurcation plane with orientation orthogonal to the trunk, respectively θ_{lb} and θ_{rb} for the left and right bifurcation plane.

As the object localization task is formulated as a classification problem, a robust detector which uses 3D Haar and Steerable Features [8] is trained using a PBT (Probabilistic Boosting Tree) [9]. In order to efficiently perform learning in high dimension space, we marginalize the search space by means of MSL [10] into subspaces which are gradually increased. Consequently, the parameter estimation problem is splitted into three steps where classifiers, parameterized by the current subset of the affine values, are sequentially trained on the subspaces: position, position-rotation and full affine transformation. A further speed improvement is achieved by using a pyramidal-based coarse-to-fine approach and searching in low-resolution (3mm) volume.

The estimated parameters θ_i initialize the landmarks $(L_t, L_{rvot}, L_{mb}, L_{lb}, L_{rb})$ by its position and orientation and constrain the search domain D_i by its scale. A further accuracy improvement is achieved by learning a discriminative 2 Level PBT classifier $\boldsymbol{H}_L$ which learns the target distribution

$$p(\boldsymbol{L}_i|x_l, y_l, z_l, \boldsymbol{I}) = \boldsymbol{H}_i(x_l, y_l, z_l|\boldsymbol{I}),\ (x_l, y_l, z_l) \in D_i,\ i = 1..5 \tag{1}$$

in 1mm volume resolution, where $p(L_i|x_l, y_l, z_l, I)$ is the probability of the presence of $\mathbf{L}_i$ at location (x_l, y_l, z_l).

3.2 Center Line Detection

In this step we initialize the center lines by previously detected landmarks. The main center line which goes through the center of the pulmonary trunk

[1] T represents discrete time steps (10 for a regular 4D cardiac CT acquisitions).

is bounded by the L_{rvot} and L_{mb}, whereas the left and right center lines are constrained by the L_{lb} and L_{rb}, respectively L_{mb} (see Section 2).

Automatic refinement of the center-line curves is achieved by applying a robust circle detector, which is trained using the PBT algorithm and Haar features. In the learning phase positive and negative samples are generated in circular form, following the annotated curves. In the detection stage, an incremental approach is used for searching 3D center points Q_i on a series of parallel planes. These are bounded by the corresponding landmarks, while their normal is given by the initial center-line and updated after each detected center point.

A least-square approach is used to fit a parametric NURBS curve C to the discrete set of detected center points Q_i. The LSE problem is solved by minimizing an objective function (Eq.2) with respect to the control points P^k

$$E = \sum_{i=1}^{N-1} |Q_i - C(\bar{u}_i)|^2, \quad \underbrace{C(u)}_{u\in[0,1]} = \frac{\sum_{k=0}^{n} N_{k,d}(u) w_k \boldsymbol{P}^k}{\sum_{k=0}^{n} N_{k,d}(u) w_k} \tag{2}$$

3.3 Dynamic Model Detection

The full surface model is initialized by projecting a mean shape, averaged over the whole annotations from the training set, into the patient specific anatomy. The projection is achieved using a piecewise affine transformation defined from the detected landmarks and center-lines.

A set of center-points C_i are equidistantly sampled along the detected center-lines. At each location, we construct a local coordinate system, using the center-line's tangent and projected RVOT-Triagone to define the orientation. The control points P^k are associated to one of the local coordinate systems, based on their Euclidean distance to the local origin. Using a coordinate system transformation we map the mean-shape control points P^k to the corresponding image location (see Fig. 2).

The transformed mean model provides accurate global fitting into the patient's specific anatomy. However, further local processing for precise object delineation is required. A boundary detector, trained by PBT in combination with steerable features, tests a set of hypotheses along the surface normal and moves the shape towards position with highest boundary probability. Spatially smooth

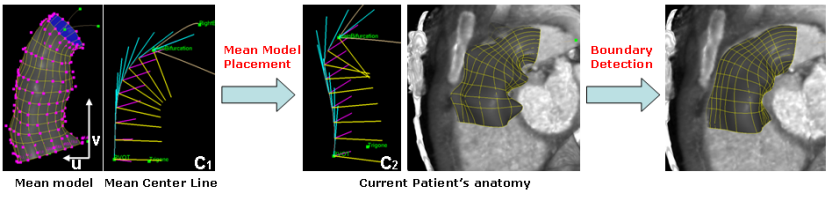


Fig. 2. Piecewise mean model placement

surface is obtained by projecting the detected surface into the corresponding shape subspace [11].

Temporal consistency of the dynamic model is enhanced similar as in [12]. One-step-forward prediction approach is applied to estimate the motion prior using motion manifold learning, which gives an appropriate initialisation for the boundary detectors. Optical flow is fused with tracking by detection to achieve temporal consistency.

4 Results

4.1 Results on Pulmonary Trunk Model Estimation

The performance of the proposed approaches is evaluated on 50 4D CT (503 volumes) studies, associate with a manual annotation considered to be the ground truth. Included data of patients with various ages (from 5 months infant to 65 years old adult) lead to significant differences in morphology and pathology of the pulmonary trunk. CT data is acquired using different protocols, resulting in volume sizes between 153x153x80 and 512x512x350 and voxel resolutions from 0.28mm to 2.0mm.

The accuracy of our detection framework is evaluated by using a three-fold cross validation. Table 1 summarizes the results from the cross-validation for the full model fitting. The detection error for the landmarks and the bounding box is measured from the Euclidean distance between the ground truth and detected landmarks, respectively bounding box corners. The point-to-mesh and point-to-curve measurement error was used for the boundary and the center lines detectors, respectively. As expected, the error constantly decreases with each estimation stage. We obtain an average precision of 2.2 mm and computation time of 202 sec on a standard 2.0GHz Dual Core PC.

4.2 Results on Clinical Evaluation

The accuracy of the automatic quantification of the pulmonary trunk is demonstrated by comparing a set of morphological and dynamic based measurements derived from the model to expert measurements and literature reported values.

Schievano et.al. [5] has proposed a set of measurements: max (RVOT) diameter, min diameter and diameter at the main bifurcation region from which the exact location for anchoring the valve stent is defined [6]. PPVI intervention

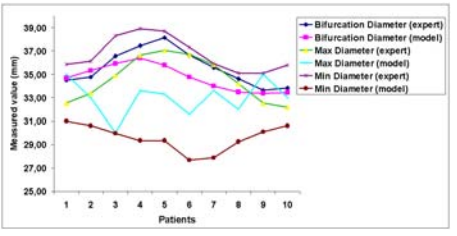
Table 1. Detection accuracy

	Mean Error(mm)	Median(mm)	Std.Dev(mm)
Bounding box	6.5	7.1	2.7
Landmarks	3.9	4.5	2.1
Center Line	4.5	4.6	1.9
Boundary	2.2	1.1	2.1

Table 2. Model-based quantification error with respect to literature(upper table) and expert measurements(lower table)

(mm)	Observed	Literature
Bifurcation Diameter	30.2 ±1.6	30.7 ±3.6

(mm)	Mean Err.	Std. Dev.
Min Diameter	1.99	0.64
Max Diameter	4.06	2.09
Bifurcation Diameter	1.04	0.65



is avoided by patients with RVOT diameter > 22mm due to device limitation [1,4,5]. Hence, the accuracy of the measurements is crucial for the success of the PPVI. The evaluation results of the proposed model-based quantification method are compared with results from the literature [13] and from experts (see Table 2).

4.3 Results on Model Based Patient Selection Suitable for PPVI

The morphology of the pulmonary trunk is a major determinant of suitability for PPVI [5]. Intervention in unsuitable patients exposes patients to unnecessary invasive catherization. A valve stent placed in such patients has a high probability of proximal device dislodgment. Schievano et.al. [5] proposed the classification of various morphologies in five groups: pyramidal shape (type I), constant diameter (type II), inverted pyramidal shape (type III), wide centrally but narrowed proximally and distally (type IV), and narrowed centrally but wide proximally and distally (type V). Patients from type I are considered to be unsuitable for PPVI due to the narrow artery and high probability of device migration. Hence, the main challenge lies in discriminating anatomies of type I from other classes.

Similar as in [14] we propose a robust shape-based patient selection for PPVI. Shape features extracted from the estimated pulmonary trunk (see Section 2) are used to learn a discriminative distance function using the Random Forest

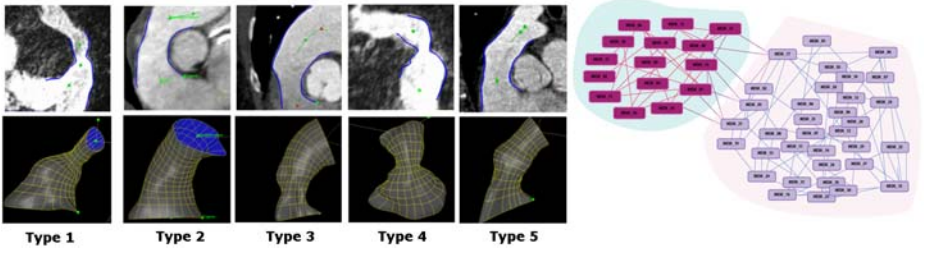


Fig. 3. Left: examples of 3D models for each morphological type. **Right**: Model-based patient classification: patient unsuitable (light cyan cluster) / suitable (light violet cluster) for PPVI intervention.

in the product space. This is applied to classify subjects into two classes: PPVI suitable and PPVI unsuitable.

The results of our method are illustrated in Fig. 3. The accuracy of the model based classification is validated by leave-one-out cross-validation. We observed correct classification in 91% of the **50** cases. The proposed method has the potential to significantly improve accuracy and reproducibility of patient selection for PPVI.

5 Discussion

This paper proposes a novel dynamic model for morphological and functional quantification of the pulmonary trunk and RVOT for PPVI intervention. Integrating all the relevant data obtained during a CT scan in a holistic fashion, the physiology of the patient's anatomy is accurately represented. An efficient and robust learning-based algorithm was proposed to compute a patient specific pulmonary model from the available image data (CT). The estimated model can be utilized to extract morphological information, direct measurements of the pulmonary trunk and the pulmonary dynamics over the cardiac cycle. Automatic model-based dynamic measurements improve the patient evaluation and selection as well as the planning of the interventional procedure. Shape-based morphology type assessment in combination with RVOT size computation provides precise selection of proper morphology appropriate for percutaneous implantable pulmonary valves and avoids unnecessary patient catheterization.

References

1. Boudjemline, Y., Agnoletti, G., Bonnet, D., Sidi, D., Bonhoeffer, P.: Percutaneous pulmonary valve replacement in a large right ventricular outflow tract: An experimental study. American College of Cardiology 43, 1082–1087 (2004)
2. Parr, J., Kirklin, J., Blackstone, E.: The early risk of re-replacement of aortic valves. The Annals of Thoracic Surgery 23(4), 319–322 (1977)
3. Carnaghan, H.: Percutaneous pulmonary valve implantation and the future of replacement. Science and Technology 20(1), 319–322 (2006)
4. Schievano, S., Migliavacca, F., Coats, S., Khambadkone, L., Carminati, M., Wilson, N., Deanfield, J., Bonhoeffer, P., Taylor, A.: Percutaneous pulmonary valve implantation based on rapid prototyping of right ventricular outflow tract and pulmonary trunk from mr data. Radiology 242(2), 490–499 (2007)
5. Schievano, S., Coats, L., Migliavacca, F., Norman, W., Frigiola, A., Deanfield, J., Bonhoeffer, P., Taylor, A.: Variations in right ventricular outflow tract morphology following repair of congenital heart disease: Implications for percutaneous pulmonary valve implantation. Journal of Cardiovascular Magnetic Resonance 9(4), 687–695 (2007)
6. Bonhoeffer, P., Boudjemline, S.A., Qureshi, Y., Bidois, J.L., Iserin, L., Acar, P., Merckx, J., Kachaner, J., Sidi, D.: Percutaneous insertion of the pulmonary valve. Journal of the American College of Cardiology 39(10), 1664–1669 (in press, 2002)
7. Piegl, L., Tiller, W.: The NURBS book. Springer, London (1995)

8. Zheng, Y., Barbu, A., et al.: Fast automatic heart chamber segmentation from 3d ct data using marginal space learning and steerable features. In: ICCV (2007)
9. Tu, Z.: Probabilistic boosting-tree: Learning discriminativemethods for classification, recognition, and clustering. In: ICCV 2005, pp. 1589–1596 (2005)
10. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Four-chamber heart modeling and automatic segmentation for 3-d cardiac ct volumes using marginal space learning and steerable features. IEEE Transactions on Medical Imaging 27(11), 1668–1681 (2008)
11. Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J.: Active shape models-their training and application. Computer Vision and Image Understanding 61(1), 38–59 (1995)
12. Yang, L., Georgescu, B., Zheng, Y., Meer, P., Comaniciu, D.: 3d ultrasound tracking of the left ventricle using one-step forward prediction and data fusion of collaborative trackers. In: CVPR (2008)
13. Nollen, G., Schijndel, K., Timmermans, J., Groenink, M., Barentsz, J., Wall, E., Stoker, J., Mulder, B.: Pulmonary artery root dilatation in marfan syndrome: quantitative assessment of an unknown criterion. Heart 87(5), 470–471 (2002)
14. Ionasec, I.I., Tsymbal, A., Vitanovski, D., Georgescu, B., Zhou, S., Navab, N., Comaniciu, D.: Shape-based diagnosis of the aortic valve. In: SPIE Medical Imaging, Orlando, USA (February 2009)

Robotic Force Stabilization for Beating Heart Intracardiac Surgery

Shelten G. Yuen[1], Michael C. Yip[1], Nikolay V. Vasilyev[2], Douglas P. Perrin[2], Pedro J. del Nido[2], and Robert D. Howe[1,3]

[1] Harvard School of Engineering and Applied Sciences, Cambridge, MA
[2] Department of Cardiovascular Surgery, Children's Hospital Boston, MA
[3] Harvard-MIT Division of Health Sciences & Technology, Cambridge, MA

Abstract. The manipulation of fast moving, delicate tissues in beating heart procedures presents a considerable challenge to surgeons. We present a new robotic force stabilization system that assists surgeons by maintaining a constant contact force with the beating heart. The system incorporates a novel, miniature uniaxial force sensor that is mounted to surgical instrumentation to measure contact forces during surgical manipulation. Using this sensor in conjunction with real-time tissue motion information derived from 3D ultrasound, we show that a force controller with feed-forward motion terms can provide safe and accurate force stabilization in an in vivo contact task against the beating mitral valve annulus. This confers a 50% reduction in force fluctuations when compared to a standard force controller and a 75% reduction in fluctuations when compared to manual attempts to maintain the same force.

1 Introduction

Beating heart surgical procedures remove the need for cardiopulmonary bypass, which has a number of well known and serious side effects for patients [1]. However, surgical manipulation inside the beating heart is challenging because heart motion exceeds the approximately one Hz human tracking bandwidth [2]. The mitral valve annulus, for instance, traverses most of its trajectory and undergoes three direction changes in approximately 100 ms, making the application of a constant, firm force for surgical tasks like mitral valve annuloplasty difficult [3].

These challenges have spurred recent interest in robotically-assisted interventions [3,4,5,6]. In prior work, we developed a 3D ultrasound-guided motion compensation system that uses an actuated, handheld robotic instrument to track beating intracardiac structures [3]. While this approach increases surgical capabilities in certain tasks [3,6], problems arise when the instrument comes into sustained contact with tissue because small positional errors can lead to significant forces and the deformation of tissue by the instrument reduces the usefulness of subsequent position measurements. In this work we consider a more direct approach by using force control of the instrument against the surgical target. We further show that feeding forward tissue motion information from 3D ultrasound can improve the performance of a force controller in this application.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 26–33, 2009.

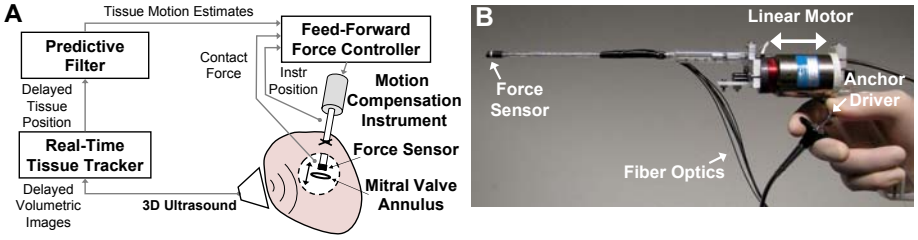


Fig. 1. (A) System overview and (B) motion compensation instrument. The instrument is controlled to a constant force against beating heart structures. Feed-forward motion estimates are derived from a 3D ultrasound tissue tracker and predictive filter.

Beating heart force control requires the development of a force sensor that can be used inside the heart. Force sensors have been developed for minimally invasive investigations [7], but they are not capable of integration with existing surgical instruments for simultaneous measurement and modification of beating heart tissue. Safety is also an important consideration when robots are used in surgery. While force control stability and performance have been studied in general settings [8], they have not been examined for beating heart applications.

In this work, we present a robotic force stabilization system for beating heart tissue manipulation (Fig. 1A). The system uses a novel, miniature uniaxial force sensor that can measure contact forces inside the heart. It is mounted to an actuated, handheld surgical instrument that we term the motion compensation instrument (Fig. 1B) for interaction with moving intracardiac structures. We describe the design of a feed-forward force controller for enhanced safety and accuracy, with feed-forward information provided by real-time 3D ultrasound tissue tracking. We validate the system in the context of beating heart mitral valve annuloplasty in an in vivo experiment in a Yorkshire pig model.

2 Force Sensor Design

A number of considerations guide the design of the force sensor for our system. First, the sensor should be located at the instrument tip to accurately measure contact forces. Second, its use inside the heart dictates that it be small, completely sealed from blood, and electrically passive to avoid disrupting conduction in the heart. Finally, to be useful in beating heart mitral annuloplasty, the sensor must be compatible with the deployment of surgical anchors.

Optically-based sensing is attractive in this setting because it does not require electrical trasmission to the sensor, has low noise, is readily miniaturized, and permits inexpensive, disposable sensors. The sensing principle relies on measuring small displacements of a reflective plate relative to the ends of optical fiber pairs (Fig. 2). Three pairs of optical fibers, with each pair comprised of one transmitting and one receiving fiber, are placed in an equilateral triangle formation at the base of the sensor to ensure that the reflective plate deflection

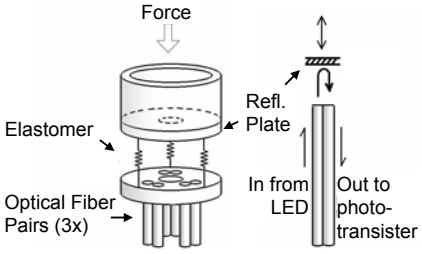


Fig. 2. Tip forces displace a reflective plate and increase light intensity to the receiving optical fibers

Fig. 3. A 14 gauge needle passes through the sensor to deploy surgical anchors

is captured entirely. An elastomer element is placed between the optical fiber ends and the reflective plate to convert force to displacement. The displacement modifies the light intensity measured by the receiving fiber, which is converted to a voltage by a phototransistor circuit.

Figs. 1B and 3 show the prototype force sensor. It is built to encompass a 14 gauge needle for the deployment of surgical anchors. Polysiloxane elastomer provides low modulus and hysteresis. The rigid housing is made of Delrin for good appearance in ultrasound images. The external diameter and length of the force sensor are 5.5 mm and 12 mm, respectively. A thin film of silicone seals the exterior surface of the sensor to shield the internal components from fluid motion. Characterization of the force sensor against a commercial sensor (ATI mini40) indicates that our sensor has a 0.17 N RMS accuracy. This was determined by applying 10 Hz bandlimited loads from 0-5 N and up to 30° incidence angle. Calibration was performed in 37° C water to match in vivo thermal conditions.

3 Force Control with Feed-Forward Motion Information

Safety is a major concern when employing robotics in surgery. In particular, a damped, stable system is desired to ensure that the robot will not oscillate in response to sudden heart motions. Unfortunately, this is at odds with maintaining fast system response to reject force disturbances. It can be difficult to meet both criteria simultaneously when small, unknown time delays are present in the system. Feed-forward target motion information can improve force controller performance against a moving target [9] and in this section we explore its use to enhance safety and performance in the context of beating heart surgery.

The motion compensation instrument may be modeled as a mass m and damper b subjected to a commanded actuator force f_a and environment contact force f_e. The damper b captures the effects of friction in the instrument, friction at the insertion point to the heart, and fluid motion. Approximating the environment as a spring of stiffness k_e yields the following system dynamics

$$m\ddot{x} + b\dot{x} = f_a - k_e(x - x_e), \tag{1}$$

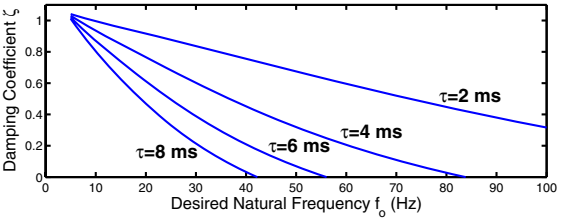


Fig. 4. Damping over desired natural frequency for small delays. $\zeta < 0$ is unstable

where x is the instrument tip position and x_e would be the position of the tissue target if it were not deformed. Now consider a standard force error negative feedback control law [8]

$$f_a = f_d + K_f(f_d - f_e) - K_v\dot{x}, \tag{2}$$

where K_f and K_v are controller gains and f_d is the desired force. For brevity, we henceforth refer to the controller of (2) simply as *force control*. Combining (1) and (2) and applying the Laplace transform gives the contact force relationship

$$F_e(s) = \frac{\frac{k_e}{m}(1+K_f)}{C(s)}F_d(s) - \frac{k_e s(s+\frac{K_v+b}{m})}{C(s)}X_e(s), \tag{3}$$

where $C(s) = s^2 + \frac{K_v+b}{m}s + \frac{k_e}{m}(1+K_f)$ is the closed-loop characteristic equation. Eq. (3) makes explicit that target motion x_e is a disturbance that perturbs f_e from f_d. Controller gains K_f and K_v must be chosen both for system stability and good rejection of x_e. The latter can be achieved by setting the natural frequency of the system much greater than the motion bandwidth of x_e. This also makes the system vulnerable to high frequency noise.

Suppose we would like to set the damping coefficient $\zeta = 1.05$ for a particular choice of the natural frequency f_o using second order system design techniques [10]. Viscoelasticity in the force sensor can result in small time delays that severely reduce the actual ζ at large f_o (Fig. 4). Noting that heart motion has significant spectral components up to at least 10 Hz [4,5,6], it is clear that it is difficult to obtain sufficient system bandwidth to safely reject x_e without precise knowledge of the time delay τ. Furthermore, empirically tuning the controller during a procedure is unfeasible for safety reasons.

An alternative strategy employs feed-forward motion information in the controller. Consider the control law

$$f_a = f_d + K_f(f_d - f_e) + K_v(\hat{\dot{x}}_e - \dot{x}) + b\hat{\dot{x}}_e + m\hat{\ddot{x}}_e, \tag{4}$$

which is (2) augmented with feed-forward estimates of the target velocity $\hat{\dot{x}}_e$ and acceleration $\hat{\ddot{x}}_e$. The contact force equation from (3) becomes

$$F_e(s) = \frac{\frac{k_e}{m}(1+K_f)}{C(s)}F_d(s) - \frac{k_e s(s+\frac{K_v+b}{m})}{C(s)}(X_e(s) - \hat{X}_e(s)).$$

Observe that the use of feedforward terms $\hat{x_e}$ and $\hat{\dot{x}_e}$ enable the cancellation of the disturbance x_e without the need to greatly increase the natural frequency of the system. The controller can then be designed with low natural frequency to avoid the effects of delay on damping and stability.

4 Tissue Motion Estimation with 3D Ultrasound

To obtain the motion terms needed in the feed-forward controller, we must first determine the position of the tissue in 3D ultrasound. To do this, we employ the real-time tissue segmentation algorithm from [3], which uses the instrument to designate a tissue target in the 3D ultrasound volumes. Fig. 5 depicts using this method to track a point on the mitral valve annulus in a beating porcine heart.

As in previous work [6], we model heart motion as a time-varying Fourier series with an offset and truncated to m harmonics

$$x_e(t) = c(t) + \sum_{i=1}^{m} r_i(t) \sin(\theta_i(t)),$$

where $c(t)$ is the offset, $r_i(t)$ are the harmonic amplitudes, and $\theta_i(t) \triangleq i \int_0^t \omega(\tau) d\tau + \phi_i(t)$, with heart rate $\omega(t)$ and harmonic phases $\phi_i(t)$. Prior to contact, measurements from the tissue tracker are used to train an extended Kalman filter to provide estimates of the model parameters $\hat{c}(t)$, $\hat{r_i}(t)$, $\hat{\omega}(t)$, and $\hat{\theta}_i(t)$. These parameters are used to generate smooth feed-forward velocity and acceleration terms for the force controller of (4) using the derivative equations

$$\hat{\dot{x}_e}(t) = \sum_{i=1}^{m} \hat{r_i}(t) i\hat{\omega}(t) \cos(\hat{\theta}_i(t)), \qquad \hat{\ddot{x}_e}(t) = -\sum_{i=1}^{m} \hat{r_i}(t)(i\hat{\omega}(t))^2 \sin(\hat{\theta}_i(t)).$$

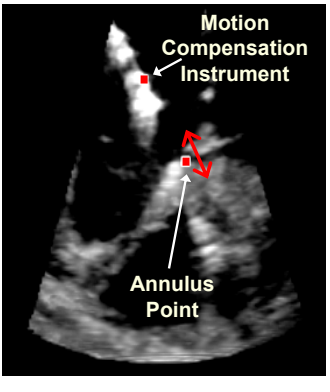


Fig. 5. Real-time 3D ultrasound tissue tracking. Squares denote instrument with force sensor and the mitral annulus.

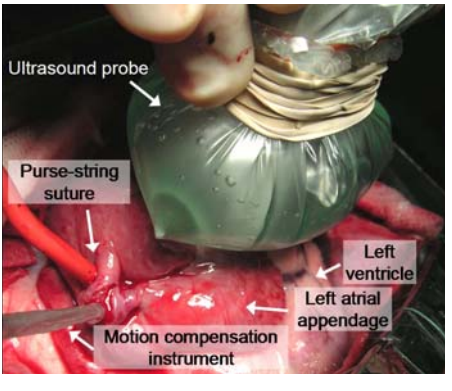


Fig. 6. In vivo experiment setup

After contact, updates to the filter are stopped because the robot interacts with the tissue, causing subsequent position measurements to no longer be representative of the feed-forward (i.e. desired) tissue motion trajectory.

5 In Vivo Validation

5.1 Experimental Setup

In vivo validation was performed in a beating Yorkshire pig heart model (Fig. 6). The tip of the motion compensation instrument was inserted into the left atrial appendage and secured by a purse-string suture. The 3D ultrasound probe (SONOS 7500, Philips Medical) was positioned epicardially on the left side of the heart to image the mitral valve and instrument. The surgeon was instructed to hold the instrument tip against the mitral annulus with a constant 2.5 N force for approximately 30 s under three conditions: manual (i.e. rigid instrument with no robot control), force control, and feed-forward force control. Controller gains were designed for $\zeta = 1.05$, $f_o = 8$ Hz based on system identification of the parameters $m = 0.27$ kg, $b = 18.0$ Ns/m, and preliminary estimates of mitral annulus stiffness $k_e = 133.0$ N/m. Contact forces were visually displayed to the surgeon during the task and recorded for offline assessment. Three trials were attempted for each condition. The experimental protocol was approved by the Children's Hospital Boston Institutional Animal Care and Use Committee.

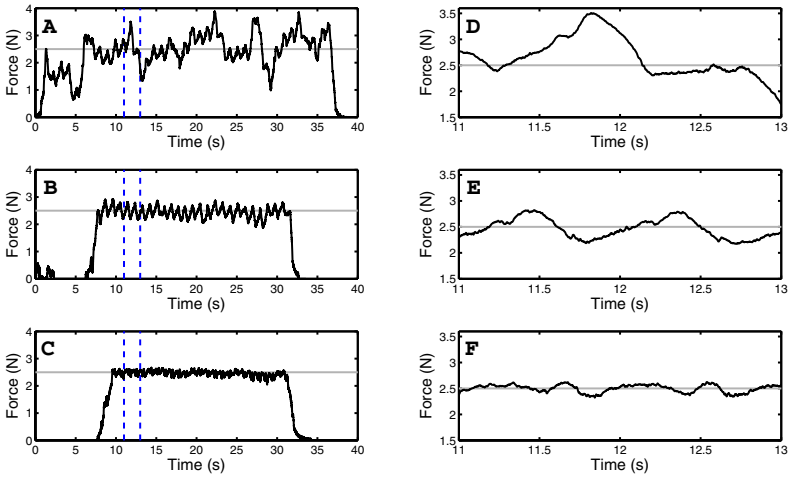


Fig. 7. Example contact forces for (A) manual, (B) force control, and (C) feed-forward force control. Corresponding enlarged views of the dashed regions in (A, B, C) are shown in (D, E, F). The desired contact force of 2.5 N is indicated (horizontal line). Data was drawn from the trials with the lowest standard deviations.

5.2 Results

Fig. 7 provides example force traces for the task executed manually, with force control, and with feed-forward force control. Averaged across all trials, manual contact with the annulus yielded force standard deviations of 0.48 ± 0.06 N (mean $\pm$ std error). Force control reduced these deviations to 0.22 ± 0.01 N with clear statistical significance in a two-sided t-test ($p = 0.012$). Feed-forward force control reduced the deviations to approximately 25% of the manual case (0.11 ± 0.02 N, $p = 0.017$). Statistical significance was also found between the force control and feed-forward control conditions ($p = 0.009$). These results are summarized in Fig. 8A. The third trial for the feed-forward force controller is omitted because the animal showed reduced viability at the end of the experiment. Performance was nearly equal to the standard force controller in this trial.

Force and feed-forward force control also reduced peak-to-peak forces (Fig. 8B). Manual use of the instrument gave swings in the contact force of 2.57 ± 0.29 N. Force control and feed-forward force control reduced these values to 1.16 ± 0.10 N and 0.65 ± 0.04 N, respectively. Statistical significance was found between all conditions at $p < 0.05$.

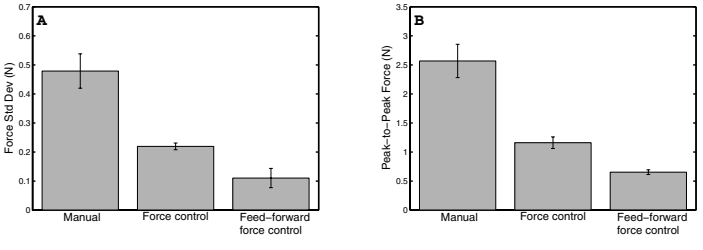


Fig. 8. Disturbance rejection measured by (A) standard deviation of forces and (B) peak-to-peak forces. Mean $\pm$ std error is shown

6 Discussion and Conclusions

In vivo results indicate that safe, accurate robotic force stabilization is feasible inside of the beating heart. The system introduced here successfully exploits near periodicity in heart motion to generate smooth estimates of tissue velocity and acceleration from noisy 3D ultrasound imaging for a feed-forward force controller. As discussed before, the use of such a controller enables conservative tuning for increased stability and damping while allowing high bandwidth force disturbance rejection. To the authors' knowledge, this is the first in vivo demonstration of force control in the beating heart.

Without robotic stabilization, placement of the instrument against the mitral annulus gave peak-to-peak force swings of 2.57 N, which is large compared to the desired 2.5 N force set point. Standard force error feedback control reduced this fluctuation by 50% and the full feed-forward controller using 3D

ultrasound motion information reduced it by another 50%. In the case of the feed-forward controller, the standard deviation of the contact forces was stabilized to 0.11 N.

In all of the force control experiments, the surgeon expressed greater confidence in instrument manipulation against the beating mitral annulus, with the feed-forward controller subjectively better than the standard force controller. These findings suggest that robotic force control may be an effective aid to the surgeon for beating heart mitral annuloplasty. We note, however, that a potential limitation of the current study is that manual tasks were done with a (nonactuated) motion compensation instrument, which is heavier than typical surgical tools. Also, an analysis of the feed-forward controller performance in conditions of noisy force measurements and motion estimates is left for future work.

While this study focused on beating mitral valve annuloplasty, we envision that the fusion of force and position sensing in a feed-forward force controller will be amenable to other beating heart procedures where there is significant and extended contact with tissue, such as catheter ablation treatment for atrial fibrillation. It may also be useful for ongoing research in robotically-assisted coronary artergy bypass graft [4,5], where small, delicate vessels are handled.

Acknowledgments. This work is supported by the US National Institutes of Health under grant NIH R01 HL073647-06.

References

1. Murkin, J., Boyd, W., Ganapathy, S., Adams, S., Peterson, R.: Beating heart surgery: why expect less central nervous system morbidity? Annals of Thoracic Surgery 68, 1498–1501 (1999)
2. Falk, V.: Manual control and tracking – a human factor analysis relevant for beating heart surgery. Annals of Thoracic Surgery 74, 624–628 (2002)
3. Yuen, S.G., Kesner, S.B., Vasilyev, N.V., del Nido, P.J., Howe, R.D.: 3D ultrasound-guided motion compensation system for beating heart mitral valve repair. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 711–719. Springer, Heidelberg (2008)
4. Ginhoux, R., Gangloff, J., de Mathelin, M., Soler, L., Sanchez, M.M.A., Marescaux, J.: Active filtering of physiological motion in robotized surgery using predictive control. IEEE Transactions on Robotics 21(1), 27–79 (2006)
5. Bebek, O., Cavusoglu, M.: Intelligent control algorithms for robotic assisted beating heart surgery. IEEE Transactions on Robotics 23(3), 468–480 (2007)
6. Yuen, S., Kettler, D.T., Novotny, P.M., Plowes, R.D., Howe, R.D.: Robotic motion compensation for beating heart intracardiac surgery. International Journal of Robotics Research (in press, 2009)
7. Puangmali, P., Liu, H., Althoefer, K.: Optical fiber sensor for soft tissue investigation during minimally invasive surgery. In: Proc. IEEE ICRA (2008)
8. Siciliano, B., Villani, L.: Robot Force Control, 1st edn. Springer, Heidelberg (1999)
9. Schutter, J.D.: Improved force control laws for advanced tracking applications. In: Proc. IEEE ICRA (1988)
10. Ogata, K.: Modern Control Engineering, 4th edn. Prentice-Hall, Englewood Cliffs (2001)

Non-rigid Reconstruction of the Beating Heart Surface for Minimally Invasive Cardiac Surgery

Mingxing Hu[1], Graeme P. Penney[2], Daniel Rueckert[3], Philip J. Edwards[4], Fernando Bello[4], Roberto Casula[5], Michael Figl[3], and David J. Hawkes[1]

[1] Centre for Medical Image Computing, University College London
[2] Department of Imaging Sciences, King's College London
[3] Department of Computing, Imperial College
[4] Department of Surgical Oncology and Technology, Imperial College
[5] Cardiothoracic Surgery, St. Mary's Hospital, London, UK
{mingxing.hu,d.hawkes}@ucl.ac.uk, graeme.penney@kcl.ac.uk,
{d.rueckert,eddie.edwards,f.bello,m.figl}@imperial.ac.uk,
roberto.casula@st-mary.nhs.uk

Abstract. This paper presents a new method to reconstruct the beating heart surface based on the non-rigid structure from motion technique using preprocessed endoscopic images. First the images captured at the same phase within each heart cycle are automatically extracted from the original image sequence to reduce the dimension of the deformation subspace. Then the remaining residual non-rigid motion is restricted to lie within a low-dimensional subspace and a probabilistic model is used to recover the 3D structure and camera motion simultaneously. Outliers are removed iteratively based on the reprojection error. Missing data are also recovered with an Expectation Maximization algorithm. As a result the camera can move around the operation scene to build a 3D surface with a wide field-of-view for intra-operative procedures. The method has been evaluated with synthetic data, heart phantom data, and in vivo data from a *da Vinci* surgical system.

1 Introduction

Minimally invasive cardiac surgery (MICS) offers great benefits to patients, such as smaller incisions and less tissue trauma, shorter hospital stays and lower infection risk. Avoiding the use of the heart-lung machine leads to less bleeding and blood loss related trauma, so the patient can return to normal life more quickly than those who have conventional open surgery. Recently telemanipulator assisted techniques have been applied to MICS. These can enhance the manual dexterity of the surgeons and enable them to more fully concentrate on the surgical procedure. Despite all these advantages, MICS still suffers from a number of problems, such as: narrow field-of-view; restricted movement; and lack of 3D guidance information.

These problems have raised concerns from both research and clinical communities and work has been done to provide 3D information from intra-operative endoscopic video [1-5]. For example, Devernay et al. addressed a multiple-step strategy to achieve 3D reconstruction of the operative scene, in which they intended to reconstruct the organ surface from stereoscopy and then fuse these small patches from each

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 34–42, 2009.

camera position to construct a larger surface [1]. Then Lau et al. published a stereo-based tracking method for cardiac surface deformation and used B-splines to model the heart surface to reduce the number of parameters to be computed [2]. Recently, Bader et al. also used stereo reconstruction to recover the 3D position of markers on a pulsating membrane to estimate the heart motion based on a state motion model [3]. However a drawback of these methods is that the images captured for 3D reconstruction are from a fixed field-of-view, so they cannot offer 3D information of peripheral scenes around the operating field. Recently Mountney et al. employed the SLAM (Simultaneous Location and Mapping) technique to build a 3D map of the scene and estimate the camera movement [4]. However, this requires long term repeatable landmarks to estimate the camera parameters, which are hard to obtain in dynamic scenes with soft tissue. More recently, Hu et al. proposed a new method to reconstruct internal organ surface based on the Structure from Motion (SFM) technique [5]. This can recover the 3D structure of the internal body from a moving endoscopic camera and so the reconstructed surface can cover a large area of the operation scene to provide a broader field-of-view for intraoperative procedures. However, this method does not allow for deformable objects and non-rigid modeling needs to be considered to improve accuracy and robustness.

This paper proposes a method to construct a heart surface from intra-operative endoscopic images of the beating heart. Our contributions are as follows. (i) The beating heart surface can be reconstructed with preprocessing using a non-rigid SFM technique. First, the beating heart cycle is robustly estimated using Hu's algorithm based on geometric constraints [6]. Then the images captured at the same position of each cycle are selected for reconstruction in order to reduce the dimensionality of the deformation subspace. Then the deformable structure is restricted to lie within a low-dimensional subspace and a probabilistic model is used to recover the 3D structure and camera motion simultaneously. (ii) The endoscopic camera can move around the operation scene. Thus, the reconstructed surface can cover a large area of the operative scene, and it can provide the surgeon with a wider field-of-view including 3D information. (iii) Outliers are removed from the feature dataset using an iterative algorithm based on the reprojection error. Outliers obtained from bad locations and false tracking can disturb the reconstruction so much that the fitting parameters become arbitrary. To address this problem we compute the reprojection error of the features after each iteration, and remove the potential outliers.

2 Methods

2.1 Pre-processing

First we calibrate the endoscope in order to have more information about the internal properties of the camera, especially the distortion parameters. Endoscopes are often designed to have barrel distortion to capture a large field of view in a single image. Barrel distortion introduces nonlinear changes to the image and would introduce large errors to the 3D reconstruction. Here we use Zhang's planar calibration [7], which requires only a few images of a planar pattern from different unknown orientations.

Because of the complexity of the cardiac motion, it is difficult to model it directly using simple deformable modeling techniques. To reduce the complexity of the heart motion for 3D reconstruction we re-arrange the endoscopic image sequence. We first analyze the heart motion based on geometric constraints using Hu's method and extract images from the same position of each heart cycle automatically for surface reconstruction. Usually we extract the images captured at the end-diastole, and for each heart cycle, two or three images are chosen. Then we employ a non-rigid modeling algorithm to model the residual cardiac motion remaining in the rearranged image sequence in order to reconstruct the heart surface. Thus we convert the difficult 4D (3D+time) dynamic reconstruction back to 3D reconstruction with small residual motion, which relatively is easier to solve. The electrocardiogram (ECG) signal could also be used to obtain the length of the heart cycle, however this requires extra hardware, an ECG machine, and synchronization between the endoscopic images and the ECG signal.

2.2 Non-rigid Structure from Motion

After the pre-processing, we employ the non-rigid structure from motion (NRSFM) [8-9] technique to reconstruct the beating heart surface using the re-arranged image sequence. Given a set of tracked feature points, the 2D and 3D locations can be related under weak-perspective projection as [10, 11]

$$\mathbf{x}_i^j = \lambda^j \mathbf{R}^j \left(\mathbf{X}_i^j + \mathbf{t}^j \right) \tag{1}$$

where $\mathbf{x}_i^j = \left(x_i^j, y_i^j \right)^T$ is the projection of 3D point $\mathbf{X}_i = \left(X_i, Y_i, Z_i \right)^T$ ($i = 1, \cdots, n$) onto the j-th ($j = 1, \cdots, \mathrm{m}$) image plane. $\mathbf{t}^j$ is a 3×1 translation vector, $\mathbf{R}^j$ is a 2×3 projection matrix, λ^j is the weak-perspective scalar. The features tracked over n frames are combined to form vectors

$$\mathbf{x}^j = \mathbf{G}^j \left(\mathbf{X}^j + \mathbf{T}^j \right) + \mathbf{N}^j \tag{2}$$

where $\mathbf{G}^j$ combines the matrix $\lambda^j \mathbf{R}^j$ and $\mathbf{N}^j$ is a noise vector added to the measurement vector $\mathbf{x}^j$. So the deformable shape can be represented in a K-dimensional linear subspace

$$\mathbf{X}^j = \overline{\mathbf{X}} + \mathbf{V}\mathbf{z}^j + \mathbf{m}^j \tag{3}$$

where $\mathbf{m}^j$ is the noise vector. Each column of the matrix $\mathbf{V}$ is a basis vector, and every entry of $\mathbf{z}^j$ is a corresponding weight that determines the contributions of the basis vector to the shape.

Torresani et al. proposed a method based on the probabilistic PCA (PPAC) model to estimate motion and learn the deformation basis [11]. The distribution of the deformable shapes is described as

$$\mathbf{x}^j \sim \mathrm{N}\left(\mathbf{G}^j \left(\overline{\mathbf{X}} + \mathbf{T}^j \right), \mathbf{G}^j \left(\mathbf{V}\mathbf{V}^T + \sigma_m^2 \mathbf{I} \right) \mathbf{G}^{jT} + \sigma^2 \mathbf{I} \right) \tag{4}$$

Then the EM algorithm is applied to estimate the PPCA model from feature points. That is, estimate the parameters $\mathbf{G}^{1:n}$, $\mathbf{T}^{1:n}$, $\overline{\mathbf{X}}$ and σ^2 from the given features $\mathbf{x}^{1:n}$. So the data likelihood can be described as

$$p\left(\mathbf{x}^{1:n} \middle| \mathbf{G}^{1:n}, \mathbf{T}^{1:n}, \overline{\mathbf{X}}, \mathbf{V}, \sigma^2 \right) = \prod p\left(\mathbf{x}^j \middle| \mathbf{G}^j, \mathbf{T}^j, \overline{\mathbf{X}}, \mathbf{V}, \sigma^2 \right)$$

The EM algorithm alternates between two steps: in the E step, a distribution over the latent coordinates $\mathbf{z}^j$ is computed; in the M step, the other variables are updated. As to the missing data $\mathbf{x}_{i'}^{j'}$ in the measurement matrix, they are also recovered during the M-step by optimizing the expected log likelihood. The reader can refer to [9] for more details.

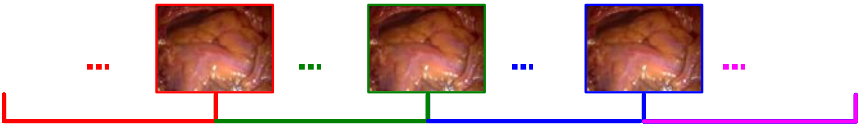


Fig. 1. Rearranged original image sequence. Images are extracted from the same phase in each beating heart cycle.

2.3 Outlier Removal

Outliers, which are caused by large errors in feature detection and tracking, are in gross disagreement with a specific postulated model and must be detected and removed from the feature database. Usually, outliers detected using epipolar geometry are features with a large geometric error (epipolar distance or Sampson distance). RANSAC (Random Sample Consensus Paradigm) is often employed as an outlier removal method [12]. However, because of the motion of the beating heart, there is still some residual motion in the rearranged image sequence obtained from Section 2.1. So the correspondences between the images do not satisfy the epipolar constraints. So rather than insisting on epipolar geometry, we use reprojection error to remove the outliers from the feature dataset.

After the estimation of the shape basis and the motion parameters, we reproject all the 3D points $\mathbf{X}_i$ ($i=1,\cdots,m$) to each image plane

$$\hat{\mathbf{x}}_i^j = \lambda^j \mathbf{R}^j \left(\mathbf{X}_i^j + \mathbf{t}^j\right) = \lambda^j \mathbf{R}^j \left(\overline{\mathbf{X}} + \mathbf{V}\mathbf{z}^j + \mathbf{t}^j\right)$$

where $\hat{\mathbf{x}}_i^j$ is the reprojection of point $\mathbf{X}_i$ on image j ($j=1,\cdots,n$). Then we construct a weight matrix $\mathbf{W}$ of size $n\times m$, in which 0s correspond to missing elements, and 1s to the others.

So the RMS error between the reprojection $\hat{\mathbf{x}}_i^j$ and the measurement $\mathbf{x}_i^j$ for point $\mathbf{X}_i$ can be denoted as

$$E_i = \sum_{j=1}^{n} w_i^j \cdot \left\| \hat{\mathbf{x}}_i^j - \mathbf{x}_i^j \right\|_F^2 \Bigg/ \sum_{j=1}^{n} w_i^j \tag{5}$$

where $\|\cdot\|_F$ is the Frobenius norm. Then the decision rule for inliers and outliers can be expressed as

$$\begin{cases} inlier & if \;\; E_i \le \gamma \\ outlier & otherwise \end{cases} \tag{6}$$

$\gamma = 1.96\sigma$ is the threshold for considering the inliers, and the standard deviation σ can be found as a maximum likelihood estimate using the median $\sigma = 1.4826\left(1+5/(n-7)\right)\sqrt{med_i \left|E_i\right|}$. An overview of the complete algorithm is given in Fig. 2

2.4 Experimental Design

In order to assess the performance of the proposed method, a beating heart phantom made of silicone rubber (The Chamberlain Group, Great Barrington, MA, USA) was employed to provide a gold standard for evaluation. The phantom is continuously inflated and deflated using an air pump with an integrated controller. In order to create a beating heart model, it was scanned while beating at the rate of 55bpm with a Philips 64-slice CT scanner, producing 10 uniformly-spaced phases.

Input: Endoscopic image sequence of the beating heart
Output: 3D shape and motion parameters

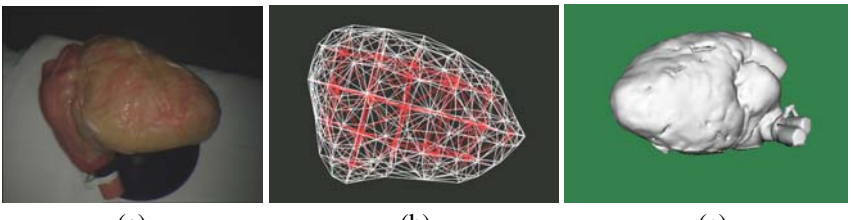
Algorithm:
Camera calibration based on Zhang's method [7]
Heart motion analysis based on Hu's method [6]
Rearrange image sequence
Feature tracking using LK tracker [13]
Feature point normalization
while $numOutlier \succ 0$
Initialization with T&K's method [10]
Non-rigid structure from motion based on EM algorithm
E-step: estimate the distribution over the latent coordinates $\mathbf{z}^j$
M-step: update the motion parameters
Outlier removal based on reprojection error
end

Fig. 2. Non-rigid reconstruction of the beating heart surface

The first of these was manually segmented and converted into a tetrahedral mesh of 709 elements and 747 degrees of freedom. The Image Registration Toolkit [14] was then used to create a sequence of 3D tensor product cubic B-splines deformations, mapping the initial mesh onto each phase in turn. Fig. 3 (a) and (b) shows the heart phantom and resulting mesh respectively. The heart phantom was also scanned without beating and a 3D model was created using the marching cubes algorithm as shown in Fig. 3 (c).

We have carried out experiments on three sets of data: synthetic test data, heart phantom data and *in vivo* data from endoscopic surgery. These sets of data and a description of the experiments are described in the following paragraphs.

In synthetic data testing, 100 points were randomly selected from the beating heart model described in Section 2.4. For simplicity, we adjust the beating heart rate to 60bpm (beats per minute). Then a moving camera was simulated with frame rate 25fps, and all these 3D dynamic points were projected to 50 image planes with different rotation and translation parameters to generate sets of 2D image points. Instead of rearranging the image sequence, we used all the images created from different positions of the heart cycle for 3D reconstruction. So the 50 images included two heart cycles.



(a) (b) (c)

Fig. 3. Heart phantom. (a) Image of the heart phantom; (b) FEM mesh from CT data of the beating phantom; (c) 3D model from CT data of the phantom without beating.

The *da Vinci*™ robotic system (Intuitive Surgical, Inc., Sunnyvale, CA, USA) was used to obtain images of the heart phantom. The beating heart rate was set to 75bpm and the endoscopic camera was moved around the scene. 250 images were captured with the frame rate 25fps and the first and last frames are shown in Fig. 5 (a) and (b) respectively.

The *in vivo* data testing consisted of endoscopic images collected from a TECAB (Totally Endoscopic Coronary Artery Bypass) surgical procedure. Fig. 6 (a) and (b) show the first and last frames respectively. The recorded ECG signal was used as ground truth for heart motion analysis, which was 19.7fpc (frame per cycle).

3 Experimental Results

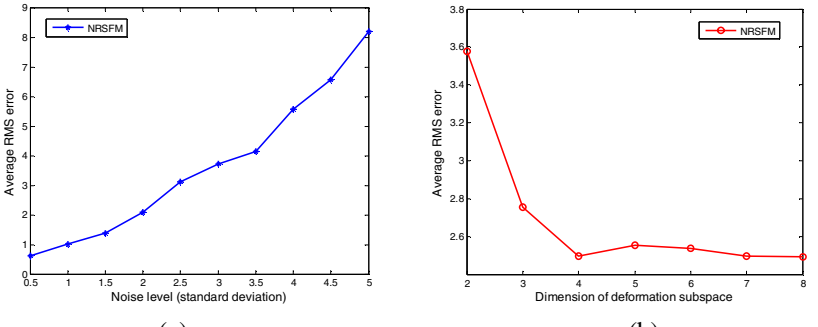
3.1 Synthetic Data

10 different ranges of Gaussian noise were added to the image measurements, with a zero mean and standard deviation varying from 0.5 to 5.0 pixels. In addition, 20% of the image points were occluded for some fraction to simulate missing data. For each noise level, we ran 200 trials and the final results were the average of results from these 200 independent experiments. A graph of these results is shown in Fig. 4 (a). It can be noticed that the RMS error of the proposed method increases gradually with the increase of noise. As the problem becomes more difficult, i.e., corrupted with severe noise ($\sigma \geq 3.5$ pixels), the error increases more quickly.

We also investigated the RMS error under different number of basis shapes as shown in Fig. 4 (b). The error decreases as the dimension K increases but it reduces more slowly when the dimension K is bigger than 4. All these indicate the proposed method can obtain an accurate result and is robust to missing data and the choice of K due to the implementation of the probabilistic PCA model and EM algorithm.

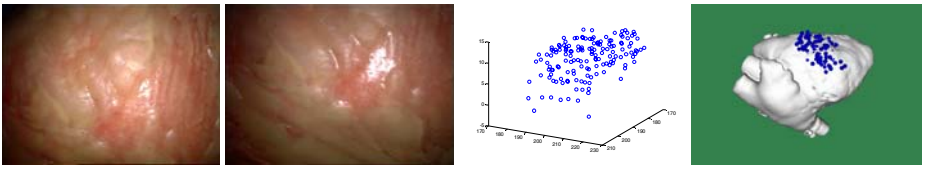
3.2 Heart Phantom Data

The first 100 images were employed to estimate the heart cycle using Hu's method [6]. The experimental result was 20.2fpc (frame per cycle), which is close to the ground truth $(60/75)*25=20$ fpc. Because the estimated result is a real value, for each heart cycle, the two images closest to the end-diastole were chosen. So 20 images were extracted



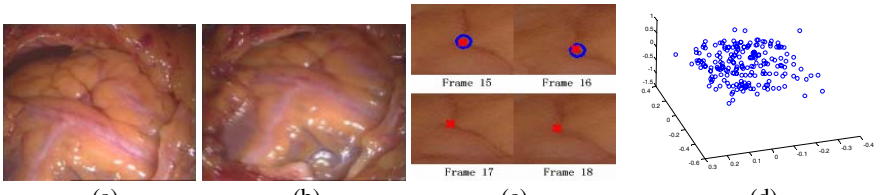
(a) (b)

Fig. 4. Experimental results of synthetic data testing. (a) Average RMS error under different noise level; (b) Average RMS error under different dimension of deformation subspace (K).



(a) (b) (c) (d)

Fig. 5. Results using the heart phantom data. (a) and (b) show the first and last frames of the image sequence of phantom heart respectively; (c) and (d) display the reconstructed point cloud and its registration result to CT model.



(a) (b) (c) (d)

Fig. 6. Experimental results of *in vivo* data testing. (a) and (b) show the first and last frames of the image sequence from TECAB surgery; (c) illustrate the data filling results of a feature; (d) displays the reconstructed point cloud using the proposed method.

from the first ten heart cycles to generate the rearranged image sequence. Then the LK feature tracker was used to track 236 feature points. The missing data rate was 27.3% because some features moved out of the scene due to the movement of the camera. Fig. 5 (c) displays the reconstructed 3D point cloud using the proposed method. Since we did not have a gold standard transformation between the 3D points and our CT image, we used the Iterative Closet Point (ICP) [15] algorithm to register the 3D points to the surface derived from CT to obtain a measure of 3D reconstruction accuracy. Fig. 5 (d) shows the final position after ICP registration. Most of the points lie on or are close to the surface and the mean residual is around 3.7 mm. Moreover, using visual inspection we can see that the positions of 3D points are overlaid on the CT surface after ICP registration in roughly the correct position.

3.3 *In vivo* Data

The estimated heart rate was 19.6fpc using Hu's method. Then 20 images were extracted from 10 heart cycles based on the estimated result. 215 features were detected and tracked from the rearranged 20 images and on average 21.1% of data in the measurement matrix was missing. Fig. 6(c) shows the results of our missing data filling algorithm on the *in vivo* data. The measured features are marked with blue "o" and the reprojected points to those image planes are marked with red "x". The feature point shown was detected and tracked until the 16th frame, however our method was able to recover its position accurately in frames 17 and 18.

Since no ground truth data was available for this patient, we report the RMS error of the surface reconstruction instead. Before removing the outliers, the average RMS error for all the visible features was 2.9mm. This reduced to 2.4mm after 16 outliers

were detected and removed from the feature dataset. Fig. 6 (d) shows the point cloud reconstructed from the endoscopic image sequence.

4 Discussion and Conclusions

This paper proposes a non-rigid reconstruction of the beating heart surface for MICS. The original image sequence is rearranged to reduce the dimension of deformable subspace of the heart surface. Then a non-rigid structure-from-motion technique is applied to recover the 3D dynamic structure and camera motion simultaneously. Outliers are removed iteratively based on the reprojection error and missing data are recovered with the EM algorithm.

Effort in the near future will focus on creating a series of 3D heart surfaces from different phases of the heart cycle to generate the whole dynamic heart surface. Incorporating a more sophisticated deformable model, e.g. a bio-mechanical model, for the non-rigid SFM is also an interesting direction. Currently it takes minutes to finish the whole estimation (pre-processing and non-rigid reconstruction), which could be a big problem to clinical utility. So we are going to use more efficient algorithms and Graphical Processing Unit (GPU) technology to achieve real-time computation. Our long term goal is to automatically reconstruct the surfaces of internal organs and register these with the preoperative data (CT or MRI) to provide more information for intra-procedural targeting, monitoring and control.

Acknowledgments. The work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant No. EP/C523016/1, and EPSRC Advanced Research Fellowship GR/T02799/02 for GPP.

References

1. Devernay, F., Mourgues, F., Coste-Maniere, E.: Towards Endoscopic Augmented Reality for Robotically Assisted Minimally Invasive Cardiac Surgery. In: Proc. International Workshop on Medical Imaging and Augmented Reality, pp. 16–20 (2001)
2. Lau, W.W., Ramey, N.A., Corso, J.J., Thakor, N.V., Hager, G.D.: Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 494–501. Springer, Heidelberg (2004)
3. Bader, T., Wiedemann, A., Roberts, K., Hanebeck, U.D.: Model-Based Motion Estimation of Elastic Surfaces for Minimally Invasive Cardiac Surgery. In: Proc. ICRA, pp. 2261–2266 (2007)
4. Mountney, P., Stoyanov, D., Davison, A.J., Yang, G.-Z.: Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 347–354. Springer, Heidelberg (2006)
5. Hu, M., Penney, G.P., Edwards, P., Figl, M., Hawkes, D.J.: 3D Reconstruction of Internal Organ Surfaces for Minimal Invasive Surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 68–77. Springer, Heidelberg (2007)

6. Hu, M., Penney, G.P., Rueckert, D., Edwards, P., Figl, M., Pratt, P., Hawkes, D.J.: A Novel Algorithm for Heart Motion Analysis Based on Geometric Constraints. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 720–728. Springer, Heidelberg (2008)
7. Zhang, Z.: A Flexible New Technique for Camera Calibration. IEEE Trans. PAMI 22(11), 1330–1334 (2000)
8. Torresani, L., Hertzmann, A., Bregler, C.: Learning Non-Rigid 3D Shape from 2D Motion. In: Proc. NIPS, pp. 1555–1562 (2004)
9. Torresani, L., Hertzmann, A., Bregler, C.: Nonrigid Structure-from-Motion: Estimating Shape and Motion with Hierarchical Priors. IEEE Trans. PAMI 30(5), 878–892 (2008)
10. Tomasi, C., Kanade, T.: Shape and Motion from Image Streams under Orthography: a Factorization Method. Int. J. Computer Vision 9(2), 137–154 (1992)
11. Bregler, C., Hertzmann, A., Biermann, H.: Recovering Non-Rigid 3D Shape from Image Streams. In: Proc. CVPR, pp. 690–696 (2000)
12. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis. Comm. ACM 24, 381–385 (1981)
13. Lucas, B., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proc. IJCAI, pp. 674–679 (1981)
14. Image registration toolkit, `http://www.doc.ic.ac.uk/~dr/software/`
15. Besl, P., McKay, N.: A Method for Registration of 3-D Shapes. IEEE Trans. PAMI 14(2), 239–256 (1992)

3D Meshless Prostate Segmentation and Registration in Image Guided Radiotherapy

Ting Chen[1], Sung Kim[2], Jinghao Zhou[2], Dimitris Metaxas[3], Gunaretnam Rajagopal[1], and Ning Yue[2]

[1] Bioinformatics Core, Cancer Institute of New Jersey, University of Medicine and Dentistry of New Jersey, New Brunswick, NJ, USA
[2] Radiation Oncology Department, Cancer Institute of New Jersey, University of Medicine and Dentistry of New Jersey, New Brunswick, NJ, USA
[3] Computer Science Department, Rutgers University, Piscataway, NJ, USA

Abstract. Image Guided Radiation Therapy (IGRT) improves radiation therapy for prostate cancer by facilitating precise radiation dose coverage of the object of interest, and minimizing dose to adjacent normal organs. In an effort to optimize IGRT, we developed a fast segmentation-registration-segmentation framework to accurately and efficiently delineate the clinically critical objects in Cone Beam CT images obtained during radiation treatment. The proposed framework started with deformable models automatically segmenting the prostate, bladder, and rectum in planning CT images. All models were built around seed points and involved in the CT image under the influence of image features using the level set formulation. The deformable models were then converted into meshless point sets and underwent a 3D non rigid registration from the planning CT to the treatment CBCT. The motion of deformable models during the registration was constrained by the global shape prior on the target surface during the deformation. The meshless formulation provided a convenient interface between deformable models and the image feature based registration method. The final registered deformable models in the CBCT domain were further refined using the interaction between objects and other available image features. The segmentation results for 15 data sets has been included in the validation study, compared with manual segmentations by a radiation oncologist. The automatic segmentation results achieved a satisfactory convergence with manual segmentations and met the speed requirement for on line IGRT.

1 Introduction

One out of every four cancers diagnosed among American men is prostate cancer, and prostate cancer is estimated to cause the death of 28,660 males in North American in 2008 [1]. Radiation Therapy (RT) has been widely accepted as an effective treatment modality for prostate cancer. RT uses high energy photon beams to deposit high doses of radiation in the prostate to kill tumor cells. However, high dose radiation can also cause short and long term toxicity to adjacent normal organs, e.g., the rectum and bladder. Intensity Modulated Radiation

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 43–50, 2009.

Therapy (IMRT) [2] is a relatively new modality that uses multiple beams and fluence modulations so that high dose is deposited to the prostate and relatively low dose to adjacent normal organs. The RT is based on the position of the prostate and surrounding organs at the time of CT simulation. Therefore, the accuracy of subsequent RT treatments is predicated on reproducing the prostate and surrounding organs' position in space at CT simulation during the daily RT treatment. The reproduction of the position is not trivial, as RT can last for eight weeks or more. Errors can be introduced due to changes in patient position, size of prostate, varying filling of the rectum or bladder. To guarantee that the prostate receives adequate dose, a margin is added to the target volume to form the Planned Target Volume (PTV), the volume to which high radiation dose is directed. Increasing the margin to the PTV, while increasing the odds of adequately treating the prostate, also adversely increases radiation dose to surrounding normal organs. Hence the need for IGRT (Image Guided Radiotherapy) [3] [4], which allows one to minimize the size of the PTV while still ensuring high dose to the prostate. One means of IGRT uses state of the art Cone Beam CT (CBCT) to acquire online treatment verification images. The main drawback of this method is that it currently requires manual segmentation and registration between the planning CT scan and the daily CBCT images, a cumbersome and time consuming task. Therefore, a fast, efficient segmentation and registration method would be extremely helpful clinically.

Many efforts have been made to develop algorithms for automatic CT and CBCT prostate segmentation [5] [6] [7]. However, interactive manual segmentations using tools with graphics user interface are still assumed to be the most reliable and time efficient means until now. Major difficulties that complicate the research in this direction include: 1) low saliency of boundary between the prostate and the adjacent organs; 2) gray scale inhomogeneity within the prostate; and 3) the visual artifacts generated by implanted seeds. A particular problem in segmenting CBCT images is that the noise-to-signal ratio is relatively high, and there is a strong demand from clinicians for high speed segmentation in order to adjust treatment plans in real time. Costa used [6] coupled deformable models to simultaneously segment the prostate and the bladder. However, the use of the Principal Component Analysis was rather empirical without a solid validation considering the relatively small testing group. Greene [8] developed a registration-segmentation framework for aligning between the planning CT and the treatment CBCT, but no explicit explanation was given regarding the segmentation method (especially for CBCT). This raises questions regarding the applicability of the method for real time IGRT data analysis. Another algorithm [9] used a Finite Element Method based framework to segment and validate prostate cancer data. The method was based on a mesh structure, which limited its flexibility when analyzing data with nonuniform deformations. We propose in this paper a segmentation-registration-segmentation framework to retrieve clinically critical objects in CT and CBCT data, in order to augment the potential of IGRT for prostate cancer. We will elaborate our methods in section 2 and present the experimental and validation results in section 3. We make our conclusions in section 4.

2 Method

In our framework, we focus on the segmentation of the prostate, bladder, and rectum. All images are pre-processed so that air regions in the rectum are masked with similar pixel values as the surrounding tissue during the registration. In CT images, models are initialized at seed points in the target organs.

2.1 Multi-object Level Set Models in CT

Three objects (prostate, bladder, and rectum) are modeled as level sets [10] using distance functions ϕ: which equals to 0 on the surface, greater than 0 inside and less than 0 outside. The intensity distribution for pixels inside each object are then calculated using nonparametric models in order to obtain the probabilistic density functions p_{ij} of arbitrary pixel j belonging to object i. We also compute the gradient map and the magnitude of gradient in the image domain. The segmentation process is of multiple steps. In each iteration, to determine the optimal classification, level set models representing different organs evolve and compete with each other for pixels. Each model evolves under the impact of three influences: 1) the inflation force based on the total gradient magnitude within the model volume; 2) the local image force at each pixel based on the probability function; and 3) the surface smoothness constraint based on the surface curvature $\nabla \cdot \frac{\nabla \phi}{\|\nabla \phi\|}$. To make the segmentation process more efficient, the model evolution and competition only take place in a narrow band around model surface, where $min(\phi_i) < \varepsilon$. We empirically choose $\varepsilon = 2$. For each pixel j in the narrow band, the value for $\phi(j)$ is recomputed using the following equation:

$$\frac{d\phi_i}{dt} = \left[\lambda_1 M_i + \nabla\lambda_2 P_i \cdot \left(\frac{\nabla\phi_i}{\|\nabla\phi_i\|}\right) - (\lambda_2 P_i + \lambda_3)\nabla \cdot \left(\frac{\nabla\phi_i}{\|\nabla\phi_i\|}\right)\right] \delta_\varepsilon(\phi_i) \qquad (1)$$

where $\delta_\varepsilon(\phi_i)$ is the regularized delta function, M_i is the inflation term derived from the gradient magnitude inside the object i, P_i is a function derived from the intensity probability competition as described in the following paragraph, and λ_3 is the weight for the smoothness on the object surface.

The competition of models are reflected by the second term on the right hand side of equation 1. The value of P is different from the probability of a pixel j belonging to object i (denoted by p_{ij}). First the value of P_{ij} is changed to $p_{ij} - \max p_{kj}$, where $k \in \{1, 2, 3\}$ and $k \neq i$. Second, at pixels with more than one $\phi_i > 0$, which equivalent to the overlapping between models, a new negative force source is introduced to locally decrease the value of both phi_i. The first modification sets the priority of models in the competition for pixels. The second modification enforces that there will be no overlapping in the segmentation result. Also for images with implanted seeds, which appear as bright spots in the image, we use local masks to block them so that $\lambda_2 = 0$ around the seeds. The segmentation results using the level set models can achieve a satisfactory accuracy. They serve as starting points for the registration process.

2.2 CT to CBCT Registration

Automating IGRT requires registration between CT and CBCT and also segmentation in each CBCT data set. In a previous research paper [8], segmentations of both the CT and the CBCT data were performed first, and then were used to guide the registration. However, we chose to analyze the data in the reverse order: registration first and the outcome is used to guide the CBCT segmentation. Major considerations behind our design are: 1) CBCT quality is typically inferior to that of CT and may deteriorate the outcome if the unreliable CBCT segmentation was used as input of registration. Therefore it is better to perform CT segmentation first, then register CT to CBCT, and finally improve the CBCT segmentation. Doing CT and CBCT segmentation and then registering would lead to error prone registration; and 2) registration provides an initialization close enough to the desired image feature, that it can speed up the CBCT segmentation process and therefore improves the overall efficiency of the IGRT.

For registration purposes, both the CT and the CBCT were re-sampled into voxels of the size of one cubic millimeter. In the experiments we presented in this paper, pelvis data were cropped around the clinically relevant volume, which included the prostate (and the surrounding PTV), the bladder, and the rectum. The segmentation results via level set models were converted into volumes in the format of binary voxel arrays. The segmented regions in CT images were assumed to be accurate since they were obtained from the CT, which has less noise and better contrast. In addition, the CT segmentation is part of the planning process, which usually does not have a time restriction, as does CBCT segmentation. Therefore the results can undergo extra improvement procedure, whether automatic or manual-based, before the registration.

The meshless model of each object is composed of sampling points on the surface of the corresponding level-set-segmented volume. During the registration, we denote the displacement at each sampling point as a combination of global motion and local deformation. To determine the global motion in the current iteration, we minimize the moving least square error at each point:

$$\sum_i w_i |f_p(u_i) - v_i| \tag{2}$$

where u_i are initial location vectors of sampling points, v_i are location vectors of sampling points in the current iteration of the registration, f_p is the global transformation functions at points of evaluation, and w_i are weights dependent on the location of the point of evaluation p. In our approach, the points of evaluation are more densely (compared to the sampling points) sampled on the surface of the segmented clinical object. We define the global motion at these points in the form of an affine transformation so that

$$f_p(u_i) = u_i \mathbf{M} + \mathbf{T} \tag{3}$$

Define $u_* = \frac{\sum_i w_i u_i}{\sum_i w_i}$ and $v_* = \frac{\sum_i w_i v_i}{\sum_i w_i}$, and also $\hat{u_i} = u_i - u_*$ and $\hat{v_i} = v_i - v_*$, we can solve $\mathbf{M}$ and $\mathbf{T}$ using

$$\mathbf{T} = v_* - u_* \mathbf{M} \tag{4}$$

$$\mathbf{M} = (\sum_i \hat{u_i}^T w_i \hat{u_i})^{-1} \sum_j w_j \hat{u_j}^T \hat{v_j} \tag{5}$$

The computational process for the global motion (i.e., **M** and **T**) is time efficient since the locations of u_i are fixed during the registration so that it is possible to pre-compute some parts in equation 5 to promptly yield the global motion. The computed global motion will be used as a global constraint to the registration. The global constraints at points of evaluation are treated same as 'demons' forces generated using local pixel gray scale. Both forces will undergo a diffusion process to redistribute to every pixels in the image to be registered. We changed the original equation of the registration force in Demons algorithm

$$\boldsymbol{u_{local}} = (m - s) \times \left(\frac{\boldsymbol{\nabla} s}{|\boldsymbol{\nabla} s|^2 + \alpha^2 (s - m)^2} + \frac{\boldsymbol{\nabla} m}{|\boldsymbol{\nabla} m|^2 + \alpha^2 (s - m)^2} \right) \tag{6}$$

by adding the global constraint in the region of the meshless model so that the global and local influences work together to guide the registration process.

$$u_{all} = \lambda_{local} u_{local} + \lambda_{global} u_{global} \tag{7}$$

where u_{global} is derived from equation 5 via the diffusion, and λ are weights. We set a threshold on the number of loops required by the registration, usually 100.

2.3 Meshless Segmentation in CBCT

The registration generated the displacement and hence the correlation between the CT and CBCT, as well as meshless models registered into the CBCT domain. To further improve the accuracy of the convergence between the registered and the reference image. The meshless models underwent another round of local deformation process. Each node of the meshless model deformed under the influence of the image features in the CBCT data, but their deformations were constrained by the internal structural constraint of each meshless model and the external constraint between models. After all model nodes arrived at equilibrium points in the CBCT domain, we computed the local deformation at pixels belonging to the volume of the model. The displacement at each pixel was calculated by minimizing the residue moving least square errors at points of evaluation. To improve the efficiency, we can use the same pre-computed kernel in the registration by using the same set of the sampling points.

$$\sum_i w_{ij} |p_j - q_i| \tag{8}$$

where q_i are the local deformations at the sampling points of the meshless model, p_j are the displacements to be decided at object surface points of evaluation j, w_{ij} are weights based on the relative distance between the pixel and model nodes. The local deformation process was equal to the segmentation of the object in the CBCT, and also helped to improve the registration between the CT and CBCT.

3 Experiments and Validation

The proposed segmentation-registration-segmentation framework has been tested on 15 data sets, each including both the planning CT and the treatment CBCT from prostate cancer patients. CT image volumes were acquired from a Philips CT scanner, while the CBCT data were acquired using a Varian On Board Imaging (OBI) system for patients treated by linear accelerators. The original pixel size for the CT data is a fixed $0.94 \times 0.94 \times 3$mm, but the original pixel size for CBCT images varies. The CT segmentation and validation take place in the original CT image domain. The registration, the CBCT segmentation, and the validation of both take place in the re-sampled image domain (1 by 1 by 1mm).

First we show the experimental results of CT segmentation. The segmentation results were compared with manual segmentations. The segmentation quality was evaluated based on the similarity between the manual segmentation and the proposed segmentation method. We focused on the volumetric evaluation, which was more relevant to the calculation of dose distribution, and was more critical in IGRT. The three measurement we used were the False Positive (FP), False Negative (FN), and True Positive (TP) of the pixels in the segmentation results, using the manual segmentation as the ground truth. The final volume similarity is computed using $S_{vol} = \frac{2TP}{2+FP-FN}$. An average of 93.2% volume similarity has been achieved for CT segmentation in 15 cases. The registration was evaluated by measuring the difference between the registered volume (from CT segmentation with manual revisions) and manual segmentation in the CBCT data. It is shown in Table. 1 that after the registration, the registered volumes retain a volumetric similarity of 91% to the manual segmentations of prostates in CBCT. For bladders and rectums, the volumetric similarities are less, but the averages are still over 85%. Some results are shown in Fig.1 and Fig.2.

The CBCT segmentation used the registration results to determine the initial location and shape of meshless deformable models. The segmentation results underwent a validation process similar to the CT segmentation. To verify the impact of the accuracy of CT segmentation on the final outcome, we used both manual and automatic CT segmentations as the input for the registration and compared the quality of the final CBCT segmentation. An average of 95% volumetric similarity was achieved using the manual CT segmentation as the input for the CT-CBCT registration. The average volumetric similarity was slightly lower at 93% when we use automatic CT segmentations.

The proposed framework is implemented on a Dell Server PE2900 with Xeon 2.66GHz CPU and 8GB RAM using Matlab. The implementation is highly time-efficient. The CT segmentation takes less than 3 minutes to segment the prostate, bladder, and rectum simultaneously. The registration of a cropped region of interest between the CT and CBCT domain takes 38 second in average for 15 cases. The CBCT segmentation takes less than one minute for the convergence between models and desired image features. The framework also needs extra time for the preprocessing of images, this usually takes another few more minutes depending on the user proficiency. Notice in IGRT there is no time restriction on

Table 1. The overall performance for prostate segmentation and registration. Column 'CT' are similarity measures of the automatic CT segmentation results against manual CT segmentations. CT Time is the time needed for CT segmentation. Column 'Registration' shows the similarity measures of the registered automatic CT segmentations against manual CBCT segmentations. Column 'CBCT' shows the similarity measure for automatic CBCT segmentations against manual CBCT segmentations.

data	CT	CT Time	Registration	Reg. Time	CBCT	CBCT Time
Set1	96%	152sec	91%	39sec	93%	57sec
Set2	92%	146sec	90%	42sec	92%	59sec
Set3	97%	137sec	94%	41sec	95%	58sec
Set4	94%	151sec	90%	35sec	92%	45sec
Set5	93%	120sec	91%	36sec	93%	48sec
Set6	91%	144sec	90%	40sec	93%	57sec
Set7	94%	141sec	92%	38sec	93%	50sec
Set8	89%	165sec	87%	37sec	89%	49sec
Set9	95%	138sec	90%	33sec	92%	42sec
Set10	90%	171sec	86%	39sec	90%	51sec
Set11	95%	152sec	92%	40sec	94%	53sec
Set12	93%	155sec	94%	36sec	95%	48sec
Set13	94%	146sec	95%	37sec	95%	47sec
Set14	95%	148sec	92%	36sec	94%	42sec
Set15	90%	160sec	91%	44sec	95%	61sec

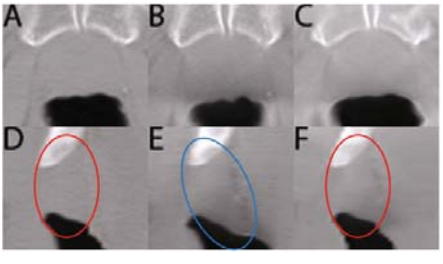


Fig. 1. Transverse and sagittal view of the registration results in the cropped image volume around the prostate. A, D are reference CT planning data; B, E are target CBCT images; and C, F shows the registered CBCT data. In D, E, F, the prostate regions are circled to show the motion in the Superior/Inferior direction.

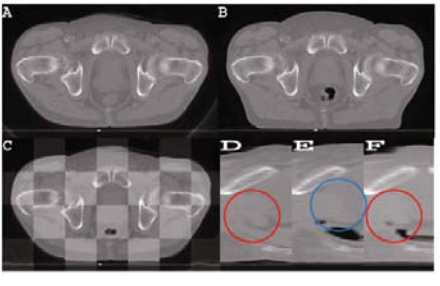


Fig. 2. Whole body registration of pelvis data. A, D are reference images; B, E are views of the target; C is a check display of the registration with registered image superimposed with the reference; and in D, E, F the prostate areas are circled out to show the motion in the Superior/Inferior direction.

CT preprocessing and segmentation. Therefore for each new treatment CBCT, with the help of previous CT segmentation, we can have clinical relevant objects registered and delineated in less than two minutes.

4 Conclusion

We have proposed a fast segmentation-registration-segmentation framework for the analysis of CBCT treatment data in IGRT. The overall accuracy of the framework is over 90% measured in shape similarity. The framework is capable of generating segmentation results in CBCT data in less than two minutes, which may inspire more IGRT applications with clinical importance.

References

1. Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T., Thun, M.J.: Cancer statistics, 2008. CA Cancer J. Clin. 58(2), 71–96 (2008)
2. Purdy, J.A.: Intensity-modulated radiation therapy. Int. J. Radiat. Oncol. Biol. Phys. 35(4), 845–846 (1996)
3. Xing, L., Cotrutz, C., Hunjan, S., Boyer, A.L., Adalsteinsson, E., Spielman, D.: Inverse planning for functional image-guided intensity-modulated radiation therapy. Phys. Med. Biol. 47(20), 3567–3578 (2002)
4. Yue, N.J., Kim, S., Jabbour, S., Narra, V., Haffty, B.G.: A strategy to objectively evaluate the necessity of correcting detected target deviations in image guided radiotherapy. Med. Phys. 34(11), 4340–4347 (2007)
5. Betrouni, N., Dewalle, A.S., Puech, P., Vermandel, M., Rousseau, J.: 3d delineation of prostate, rectum and bladder on mr images. Comput. Med. Imaging Graph 32(7), 622–630 (2008)
6. Costa, M.J., Delingette, H., Novellas, S., Ayache, N.: Automatic segmentation of bladder and prostate using coupled 3d deformable models. In: Med. Image Comput. Comput. Assist. Interv. Int. Conf., vol. 10(pt. 1), pp. 252–260 (2007)
7. Davis, B.C., Foskey, M., Rosenman, J., Goyal, L., Chang, S., Joshi, S.: Automatic segmentation of intra-treatment ct images for adaptive radiation therapy of the prostate. In: Med. Image Comput. Comput. Assist. Interv. Int. Conf., vol. 8(Pt 1), pp. 442–450 (2005)
8. Greene, W.H., Chelikani, S., Purushothaman, K., Chen, Z., Knisely, J.P., Staib, L.H., Papademetris, X., Duncan, J.: A constrained non-rigid registration algorithm for use in prostate image-guided radiotherapy. In: Med. Image Comput. Comput. Assist. Interv. Int. Conf., vol. 11(Pt 1), pp. 780–788 (2008)
9. Brock, K.K., Nichol, A.M., Menard, C., Moseley, J.L., Warde, P.R., Catton, C.N., Jaffray, D.A.: Accuracy and sensitivity of finite element model-based deformable registration of the prostate. Med. Phys. 35(9), 4019–4025 (2008)
10. Malladi, R., Sethian, J.A.: Image processing via level set curvature flow. Proc. Natl. Acad. Sci. U S A 92(15), 7046–7050 (1995)

A Computer Model of Soft Tissue Interaction with a Surgical Aspirator

Vincent Mora, Di Jiang, Rupert Brooks, and Sébastien Delorme

Industrial Materials Institute - National Research Council, Canada
vincent.mora@imi-cnrc-nrc.gc.ca

Abstract. Surgical aspirators are one of the most frequently used neurosurgical tools. Effective training on a neurosurgery simulator requires a visually and haptically realistic rendering of surgical aspiration. However, there is little published data on mechanical interaction between soft biological tissues and surgical aspirators. In this study an experimental setup for measuring tissue response is described and results on calf brain and a range of phantom materials are presented. Local graphical and haptic models are proposed. They are simple enough for real-time application, and closely match the observed tissue response. Tissue resection (cutting) with suction is simulated using a volume sculpting approach. A simulation of suction is presented as a demonstration of the effectiveness of the approach.

1 Introduction

Rapidly evolving surgical techniques, patient safety concerns, and the inherent inefficiency of operating room training are strongly driving the need for innovative simulation technologies [1,2,3]. Clinical adoption of virtual reality simulation would result in accelerated training, rapid adoption of new techniques, better surgeries with minimal risk and consequently improved patient care [4]. Our research project aims to develop a simulator capable of training medical students to perform surgical resection of brain tumors.

As a fundamental surgical device in neurosurgery [5,6,7], the surgical aspirator must be accurately modelled. This tool has two main functions: (A) aspiration, which is either the non-traumatic removal of blood and fluid or the removal of soft tissue [8,9], and (B) tissue holding [5]. Surgical aspirators are included in commercial simulators but studies on their mechanical behavior are scarce in the literature and do not provide enough experimental data to develop a model suitable for a simulator [10,11,12,13,14,15,16,17].

A surgical aspirator model is suitable for a simulator if it allows to perform functions (A) and (B) in real time [18]. The objective of this study is (1) to measure the interaction between the surgical aspirator and brain tissue, (2) to propose a haptic and graphic model of aspirator tissue interaction based on experimental evidence, and (3) to implement an algorithm for tissue holding and removal with a virtual aspirator. Fluid removal is out of the scope of this paper. As a secondary objective potential phantom material were also experimentally tested in order to find a substitute for brain tissue mechanical testing.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 51–58, 2009.

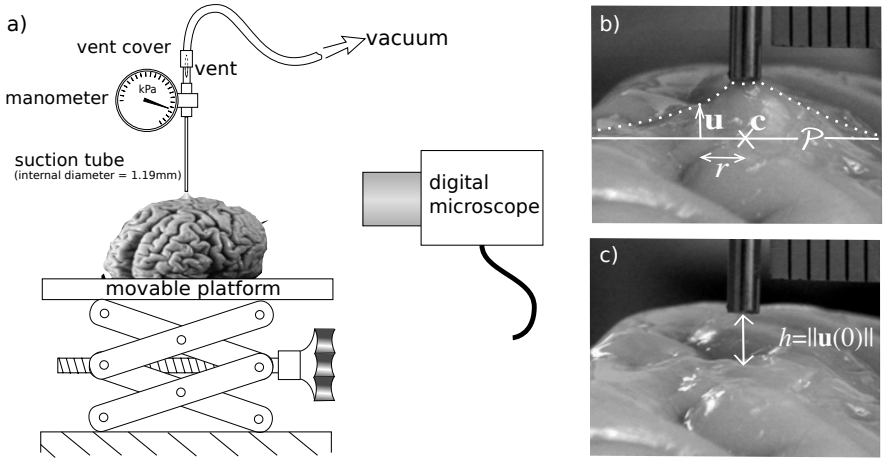


Fig. 1. a) Overview of the experimental setup. b) Photograph taken by the digital microscope before the sample detaches in order to determine the local deformation (white dotted line). Note the ruler used for calibration on the upper right corner. c) Photograph taken by the digital microscope just after the sample detached from the suction tube.

2 Experimentation

One calf brain was collected at a slaughterhouse, transported in PBS (phosphate buffered saline) and frozen at −80°C. It was then thawed 24h prior to the experiment. The tests were carried out on the surface of the calf brain covered by the pia matter with and without moisturizing. The pia matter was removed and the test with moisturizing was repeated. Two phantom materials have been tested: dessert gelatine diluted in boiling water with mass concentration of 5%, 10% and 15%; and two different brands of soft tofu.

The experimental setup, shown in figure 1 a, consisted of a movable platform, a digital microscope and a 1.19mm diameter suction tube connected to a vacuum pump. The vacuum hose featured a vent with a sliding cover to vary the suction pressure and a manometer capable of recording negative pressures between 4kPa and 80kpa with a 1kPa precision at the tip of the tube. A ruler was used for photogrammetric calibration, as shown on the upper right corner of figure 1 b and c.

The experiment proceeded as follows:

1. The material sample was put on the movable platform (see figure 1 a).
2. The vent was totally opened.
3. The sample was raised until it contacted the suction tube. The pressure was recorded when the sample completely blocked the tube oppening.
4. The platform was lowered and photographs were taken at 1mm intervals in order to record the shape of the surface in the vicinity of the suction tube.

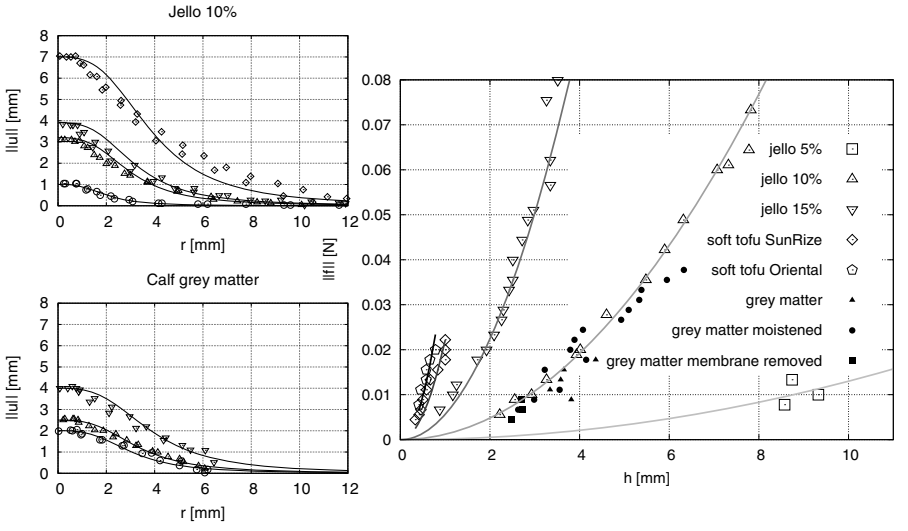


Fig. 2. Left: profile for different value of h along with fits with the function $||\boldsymbol{u}(r)|| = h^2/(h + \alpha r^3)$ (see eq. 2) with $\alpha_{\text{grey matter}} = 0.07$ and $\alpha_{\text{gelatine 10\%}} = 0.12$. Right: Force as a function of the vertical displacement h for different materials. The solid lines are fits with the quadratic function βh^2 (see eq. 1) with $\beta_{\text{gelatine 5\%}} = 1.3 \times 10^{-4}$, $\beta_{\text{gelatine 10\%}} = \beta_{\text{grey matter}} = 1.2 \times 10^{-3}$, $\beta_{\text{gelatine 15\%}} = 5.6 \times 10^{-3}$, $\beta_{\text{tofu A}} = 2.2 \times 10^{-2}$ and $\beta_{\text{tofu B}} = 3.8 \times 10^{-2}$.

5. When the sample detached from the suction tube:
   (a) If the surface of the sample was undamaged, a digital photograph was taken. The vent was then further obstructed to increase the negative pressure at the tool tip, and the process went back to step 3.
   (b) If the surface of the sample was damaged, the rupture pressure was recorded and the experiment ended.

Profiles obtained for gelatine 10% and moisturized calf grey matter without protective membrane are shown in figure 2. The distance h between the surface of the sample and the extremity of the suction tube was measured on the photographs taken just after the sample detached from the tube (see figure 1 c). Knowing the internal diameter of the suction tube d, the pressure p, and assuming frictionless contact between the tube and the sample, the magnitude of the force $\boldsymbol{f}$ when the sample detaches from the tube is given by $||\boldsymbol{f}|| = \frac{1}{4}\pi d^2 p$. The experimental data obtained for various materials are shown in figure 2. It can be observed that the calf grey matter has the same behavior as the gelatine 10% solution with a lower rupture force (38mN $\equiv$ 34kPa) that is further lowered by absence of moisturizing (18mN $\equiv$ 16 kPa) or the removal of the protective membrane (9mN $\equiv$ 8kPa).

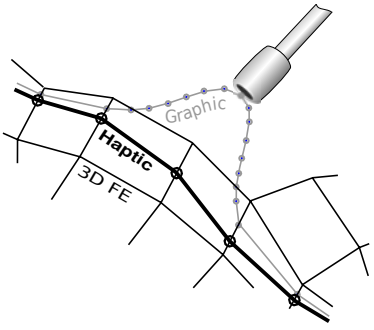


Fig. 3. The three different meshes used in the simulation

3 Simulating a Surgical Aspirator

The deformation caused by suction on tissue is both localized and large. It has a typical shape which is important to render to achieve realism in the simulation. Since the phenomenon is localised, a mixed model has been chosen. The global displacement and force fields are given by a 3D explicit non-linear finite element (FE) model[1] the details of which are outside the scope of this article (see e.g. [19,20] for representative examples). The models for the local displacement and force fields are discussed below.

Tissue deformation can be simulated using FE models. However a FE model would run too slowly to be used for haptics or detailed graphic rendering. Therefore the use of three different models and corresponding meshes is proposed, based on experimental results presented in section 2:

1. A 3D FE mesh is used for the deformation calculation.
2. A surface mesh which is a tesselation of an isosurface of a signed distance field defined on the FE mesh, is used for collision detection and haptics rendering.
3. A locally refined copy of the surface mesh is used for graphics rendering.

These three meshes are illustrated in figure 3. A overview of the complete suction simulation is presented in algorithm 1.

3.1 Tissue Holding

At lower pressures, when close enough to the tissue, tissue is attracted and sticks to the tip of the surgical aspirator. This exerts a force on the tissue, and may be used to hold, or manipulate it. The force $\boldsymbol{f}$ sent to the haptic device, when at distance h from local plane $\mathcal{P}$ (see figure 1), can be expressed as:

$$\boldsymbol{f}(h) = -\beta h^2 \boldsymbol{n}$$

[1] Modelling the local surgical aspirator effect directly using a 3D FE model would require a mesh of extraordinary detail, which would not run in real time.

Algorithm 1. Providing haptic and visual feedback

Input: Haptic device position
Result: Visual and haptic feedback
while *simulation is not over* **do**
  calculate the deformation of the FE mesh and update postion of surface mesh;
  if *grabbed vertices is empty* **then**
    grab surface mesh vertices in the vicinity of the tool tip $\boldsymbol{t}$;
  else
    compute $\mathcal{P}$: average position and normal of grabbed vertices;
    compute haptic force $\boldsymbol{f} = -\beta h^2 \boldsymbol{n}$;
    if $||\boldsymbol{f}|| < ||\boldsymbol{f}||_{\text{rupture}}$ **then**
      apply force $\frac{\boldsymbol{f}}{n}$ to the FE mesh grabbed vertices ;
      locally refine and deform graphic mesh with $\boldsymbol{u}(r) = \frac{h^2}{h+\alpha r^3}\boldsymbol{n}$;
    else
      cut FE mesh and extract new surface mesh;
    end
    empty grabbed vertices list ;
    grab surface mesh vertices in the vicinity of the projection $\boldsymbol{c}$ of $\boldsymbol{t}$;
  end
end

where β is a material parameter that increases with the stiffness. This model shows good agreement with experimental data in figure 2. In addition to being rendered by the haptic device, the force can be applied to the 3D FE model to calculate a deformation. When $||\boldsymbol{f}||$ is smaller than the experimentally measured value $||\boldsymbol{f}||_{\text{rupture}}$, tissue will be held by the aspirator. Beyond this critical value, the tissue breaks, and is removed by the aspirator as described in the next section.

3.2 Tissue Removal

To enable modelling of tissue cutting while avoiding large changes to the 3D FE mesh a volume sculpting approach [21] is used. The boundaries of soft tissues are modelled as the zero isosurface of a distance field, $F(\boldsymbol{x})$, defined on the nodes of the 3D FE mesh. Tissue removal is modelled by changing the value of F based on the position of the surgical aspirator. A similar approach has been used to simulate cutting of the petrous bone in [22].

If the aspiration force is greater or equal than $||\boldsymbol{f}||_{\text{rupture}}$, a spherical cutting region around the tool tip, $\boldsymbol{t}$, becomes active. Tissue within this sphere is removed. To do this, the distance field is updated according to:

$$F_{\text{new}}(\boldsymbol{x}) = \min\left(F_{\text{old}}(\boldsymbol{x}), ||\boldsymbol{x} - \boldsymbol{t}|| - R\right) \tag{1}$$

where R is the radius of the cutting sphere.

Once this function is changed, it is necessary to tesselate the new zero isosurface. This is done using one of a family of algorithms, which we refer to as

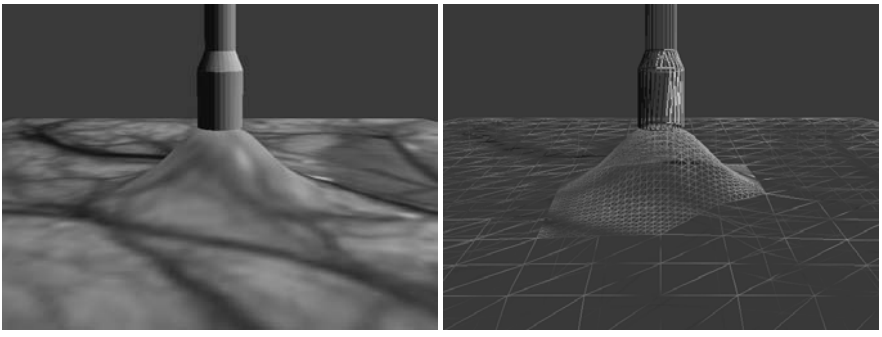


Fig. 4. Simulation of suction effect on tissue using a local rendering model. Left: screenshot from simulation; Right: same showing refined mesh near the surgical aspirator.

marching shapes [23], which tesselate each element given the field values at their vertices. The generated surface consist of triangles, whose vertices lie on the edges of the FE mesh. The most widely known such algorithm is the marching cubes of Lorensen and Cline [24], but the same approach can be used on other shapes, such as tetrahedra, octahedra, etc.

3.3 Graphics Rendering

The displacement of a point $\boldsymbol{x}$ on the brain surface due to suction is modelled by the function:

$$\boldsymbol{u}(r) = \frac{h^2}{h + \alpha r^3}\boldsymbol{n} \tag{2}$$

where α is a material parameter, h is the distance between the tool tip $\boldsymbol{t}$ and the local plane $\mathcal{P}$, r is the distance between $\boldsymbol{x}$ and the projection $\boldsymbol{c}$ of $\boldsymbol{t}$ on $\mathcal{P}$, and $\boldsymbol{n}$ is the normal to $\mathcal{P}$ (see figure 1 b). Figure 2 shows good agreement of the model with the experimental data.

The triangle elements of the surface mesh that are close to the suction tool are recursively subdivided into smaller triangles. All the nodes of the mesh are then displaced according to the model. Figure 4 shows a screenshot of the simulated interaction between surgical aspirator and brain tissue with a surface mesh that is locally refined for graphics rendering. When cutting is taking place, the use of this graphic model gives the appearance that the interior parts of the tissue pops up into the aspiration tool.

4 Discussion

In this paper, we have described a model to simulate the effect of a surgical aspirator on tissue. By separating the haptic and graphical models from the FE model, the experimentally observed haptic and visual effects of aspirating tissue were simulated in a computationally efficient manner.

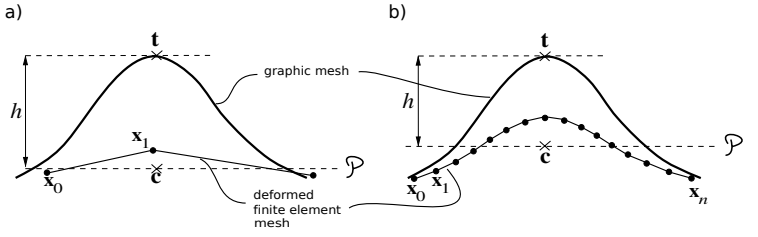


Fig. 5. Definition of $\mathcal{P}$ and h when the 3D FE mesh is a) coarse, b) fine

The experiments with phantom material suggest that gelatine (10% mass concentration) could be an adequate substitute for brain for mechanical testing purposes. This preliminary result must be validated using fresh brain tissue since freezing can alter the properties of biological tissues [25]. Other work [26] has shown that these effects can be compensated by applying a corrective factor. Thus we think it likely that with further work a suitable concentration of gelatine can be found to mimic the properties of fresh brain.

The proposed model is limited by the dependency of the definition of the local plane $\mathcal{P}$ on the refinement of the 3D FE mesh (see figure 5). In the experiment $\mathcal{P}$ (figure 1) is the average plane of the undeformed brain surface. In the model $\mathcal{P}$ (algorithm 1) is the average plane of the surface of the deformed 3D FE mesh. If the 3D FE mesh is coarse, compared to the local deformation caused by the aspirator, the local deformation of the 3D FE mesh is small and the model $\mathcal{P}$ is close to the experimental $\mathcal{P}$. However if the 3D FE mesh is fine, its local deformation will cause the model $\mathcal{P}$ to diverge from the experimental $\mathcal{P}$ which affects the calculation of $\boldsymbol{f}(h)$. This problem is not a real concern in our case since if the finite element mesh can be made fine enough for this problem to be significant, the mesh would then be fine enough to avoid the need for this local model altogether.

Future work will focus on modelling the interaction of other neurosurgical tools such as the Cavitron Ultrasonic Surgical Aspirator (CUSA), as well as on validating the proposed model against a FE simulation results obtained with a fine mesh, and against experimental data obtained on fresh human brain tissue.

References

1. Vuskovic, V., Kauer, M., Szekely, G., Reidy, M.: Realistic force feedback for virtual reality based diagnostic surgery simulators. In: Proceedings of the ICRA (2000)
2. Spicer, M.A., Apuzzo, M.L.: Virtual reality surgery: neurosurgery and the contemporary landscape. Neurosurgery 52(3), 489–498 (2003)
3. Rosenthal, R., et al.: The future of patient safety: surgical trainees accept virtual reality as a new training tool. Patient Safety in Surgery 2 (2008)
4. Mauro, A.D., Raczkowsky, J., Wirtz, R., Wörn, H.: Development of a microscope embedded training system for neurosurgery. Biomedical Simulation, 209–214 (2008)
5. Yasargil, M.G.: Microneurosurgery: microsurgery of CNS tumors. Thieme (1996)
6. Cuatico, W.: Neurosurgical suction tips. Acta Neurochirurgica 46, 303–306 (1979)

7. Vällfors, B.: Neurosurgical suction systems, an experimental study. Thesis, Göteborg, Sweden (1976)
8. Vällfors, B.: Efficient atraumatic liquid suction by means of slit suction tubes combined with a pressure control unit. Neurosurgical Review 7, 179–183 (1984)
9. Maxwell, R.E., Chou, S.N.: Convexity meningiomas and general principles of meningioma surgery. In: Operative Neurosurgical Techniques: Indications and Methods, pp. 491–501. Grune and Stratton, New York (1982)
10. Papadopoulos, S.M., McGillicuddy, J.E.: A simple device for controlling suction force in neurosurgical operations. Neurosurgery 25, 662–663 (1989)
11. Johnson, J.P., Becker, D.P.: A continuous microneurosurgical irrigation and suction system: technical note. Neurosurgery 39, 409–411 (1996)
12. Shalit, M.N.: A simple method for the control of suction power during neurosurgery. Acta Neurochirurgica 64, 235–236 (1982)
13. Lefranc, F., Brotchi, J.: Performance of a new type of suction tip attachment during intramedullary tumor dissection: technical note. Neurosurgery 61 (2007)
14. Aoki, T., Ohashi, T., Matsumoto, T., Sato, M.: The pipette aspiration applied to the local stiffness measurement of soft tissues. Annals of Biomedical Engineering 25, 581–587 (1997)
15. Vällfors, B.: Suction in neurosurgery. Acta Neurochirurgica 55, 35–42 (1980)
16. Mazza, E., Grau, P., Hollenstein, M., Bajka, M.: Constitutive modeling of human liver based on in vivo measurements. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 726–733. Springer, Heidelberg (2008)
17. Schiavone, P., Boudou, T., Promayon, E., Perrier, P., Payan, Y.: A light sterilizable pipette device for the in vivo estimation of human soft tissues constitutive laws. In: Proceedings of the 30th EMBC, pp. 4298–4301 (2008)
18. Berkley, J., Weghorst, S., Gladstone, H., Raugi, G., Berg, D., Ganter, M.: Fast finite element modeling for surgical simulation. In: Medicine Meets Virtual Reality, pp. 55–61. IOS Press, Amsterdam (1999)
19. Taylor, Z.A., Cheng, M., Ourselin, S.: High-speed nonlinear finite element analysis for surgical simulation using graphics processing units. Transactions on Medical Imaging 27(5), 650–662 (2008)
20. Miller, K., Joldes, G., Lance, D., Wittek, A.: Total lagrangian explicit dynamics finite element algorithm for computing soft tissue deformation. Communications in Numerical Methods in Engineering 23, 121–134 (2007)
21. Galyean, T.A., Hughes, J.F.: Sculpting: an interactive volumetric modeling technique. Computer Graphics 25(4), 267–274 (1991)
22. Pflesser, B., Petersik, A., Tiede, U., Hohne, K.H., Leuwer, R.: Volume cutting for virtual petrous bone surgery. Computer Aided Surgery 7(2), 74–83 (2002)
23. Newman, T.S., Yi, H.: A survey of the marching cubes algorithm. Computers and Graphics 30, 854–879 (2006)
24. Lorensen, W., Cline, H.: Marching cubes: a high resolution 3d surface construction algorithm. Computer Graphics 21(4), 163–169 (1987)
25. Fallenstein, G.T., Hulce, V.D., Melvin, J.W.: Dynamic mechanical properties of human brain tissue. Journal of Biomechanics 2(3), 217–226 (1969)
26. Lippert, S., Grimm, M.: Estimating the material properties of brain tissue at impact frequencies: A curve-fitting solution. In: Summer Bioengineering Conference, Key Biscayne, Florida (2003)

Optimal Matching for Prostate Brachytherapy Seed Localization with Dimension Reduction*

Junghoon Lee[1], Christian Labat[2], Ameet K. Jain[3], Danny Y. Song[4], Everette C. Burdette[5], Gabor Fichtinger[2,6], and Jerry L. Prince[1]

[1] Department of Electrical and Computer Eng., Johns Hopkins University, USA
[2] Department of Computer Science, Johns Hopkins University, USA
[3] Philips Research North America, USA
[4] Department of Radiation Oncology, Johns Hopkins School of Medicine, USA
[5] Acoustic MedSystems, Inc., USA
[6] School of Computing, Queen's University, Canada

Abstract. In prostate brachytherapy, x-ray fluoroscopy has been used for intra-operative dosimetry to provide qualitative assessment of implant quality. More recent developments have made possible 3D localization of the implanted radioactive seeds. This is usually modeled as an assignment problem and solved by resolving the correspondence of seeds. It is, however, NP-hard, and the problem is even harder in practice due to the significant number of hidden seeds. In this paper, we propose an algorithm that can find an optimal solution from multiple projection images with hidden seeds. It solves an equivalent problem with reduced dimensional complexity, thus allowing us to find an optimal solution in polynomial time. Simulation results show the robustness of the algorithm. It was validated on 5 phantom and 18 patient datasets, successfully localizing the seeds with detection rate of ≥ 97.6 % and reconstruction error of ≤ 1.2 mm. This is considered to be clinically excellent performance.

1 Introduction

Low dose rate permanent brachytherapy is widely utilized for low risk prostate cancer, the success of which mainly depends on the ability to place an adequate number (50–120) of radioactive seeds to deliver a sufficient therapeutic dose distribution to the target gland while sparing adjacent organs from excessive radiation. During the procedure, the surgeon implants radioactive seeds based on a pre-operative implantation plan under transrectal ultrasound image guidance. However, it is not possible to accurately implant the seeds to the preplanned positions due to various reasons, e.g., patient motion, needle deflection within prostate, and edema. In order to improve outcomes and reduce complications, intra-operative localization of the seeds using x-ray fluoroscopy and intra-operative dosimetry modifications have been previously proposed [1,2,3].

The seed localization problem from multiple fluoroscopy images are usually modeled as an assignment problem [3]. This approach resolves which segmented

* This work has been supported by DoD PC050042 and NIH/NCI 2R44CA099374.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 59–66, 2009.

seeds in each projection image correspond to the same physical seed. The 3D locations of the seeds are determined by computing so-called symbolic intersection of the lines connecting the segmented seeds to the x-ray source positions based on the revealed seed correspondence. However, this assignment problem is NP-hard [3]. In addition, there exist a significant number of hidden seeds in every image, thus making explicit segmentation of seeds in every image hard. Such so-called hidden seeds are usually determined manually, and it is sometimes impossible to recover them when one seed completely hides another. Therefore, an algorithm that is computationally efficient and is able to solve the hidden seed problem is essential to a clinically feasible system.

There has been some research on solving the hidden seed problem. Su *et al.* [4] extended Fast-CARS [2], but the new algorithm was prone to reconstruct a greater number of seeds than were actually present, an effect called "false positive" seeds. Narayanan *et al.* [5] proposed a pseudo-seed-matching strategy coupled with an epipolar geometry-based reconstruction. This method requires at least one of the three images to be complete, however, or it may or may not reliably reconstruct the 3-D seed positions of the hidden seeds. Su *et al.* [6] proposed an adaptive grouping technique which divides the seed images into groups for efficient seed reconstruction and solving the hidden seed problem. Unfortunately, it may fail to detect overlapping seeds when the projection with the largest number of seed images among the divided groups is incomplete. Also, incorrect division of triplets may result in false positive seeds. Tomosynthesis [7] and Hough trajectory [8] methods have also been proposed, but they require unfeasibly large numbers of images in order to guarantee stable reconstruction.

We have previously proposed a dimension reduction approach for solving seed matching problem [9]. However, it did not solve the hidden seed problem and was not computationally efficient. In this paper, we present an algorithm that is able to solve the hidden seed problem using dimensionality reduction to achieve efficient cost computation. We also propose a pruning algorithm that yields a dramatic reduction in computation time.

2 Method

2.1 Extended Assignment Problem

When at least three projection images are used and all the 2D seed locations are identified in every x-ray image, the correspondence problem can be formulated as a 3D assignment problem (3DAP) [3]. In reality, however, there are a significant number of hidden seeds, resulting in a varying number of segmented seeds in each image. Here, we describe an extended assignment problem (EAP) that is able to reconstruct seed positions including hidden seeds.

In contrast to the 3DAP where exactly N implanted seeds are identified in every image, we consider a different number N_i of identified seeds in each image i with $N_i \leq N$. For $I(\geq 3)$ x-ray images, the EAP is defined as:

$$\min_{x_{i_1 i_2 i_3 \ldots i_I}} \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \sum_{i_3=1}^{N_3} \cdots \sum_{i_I=1}^{N_I} c_{i_1 i_2 i_3 \ldots i_I} x_{i_1 i_2 i_3 \ldots i_I}, \tag{1}$$

$$\text{s.t.} \begin{cases} \sum_{i_2=1}^{N_2} \sum_{i_3=1}^{N_3} \cdots \sum_{i_I=1}^{N_I} x_{i_1 i_2 i_3 \ldots i_I} \geq 1, & \forall i_1 \\ \vdots & \\ \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \cdots \sum_{i_{I-1}=1}^{N_{I-1}} x_{i_1 i_2 i_3 \ldots i_I} \geq 1, & \forall i_I \\ \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \sum_{i_3=1}^{N_3} \cdots \sum_{i_I=1}^{N_I} x_{i_1 i_2 i_3 \ldots i_I} = N & \end{cases} \tag{2}$$

where $c_{i_1 i_2 i_3 \ldots i_I}$ is the cost of matching point $p_{i_1}^1$ to points $\{p_{i_2}^2, p_{i_3}^3, \ldots, p_{i_I}^I\}$ and $x_{i_1 i_2 i_3 \ldots i_I}$ is a binary variable deciding the correctness of the match $\langle i_1, i_2, \ldots, i_I \rangle$. Since we use inequalities in the constraints (2) to handle the occurrence of hidden seeds, a point can be chosen more than once in an image. The last equality forces the total number of seeds to be N.

Let $N' = N_1 N_2 \ldots N_I$, and $\boldsymbol{x}, \boldsymbol{c} \in \mathbb{R}^{N'}$ be vector forms of $x_{i_1 i_2 \ldots i_I}$ and $c_{i_1 i_2 \ldots i_I}$, respectively. Let $\mathbf{M}$ be a matrix form of (2) except for the last equation. Then the EAP (1)–(2) can be formulated as the following integer program:

$$P: \quad \min_{\boldsymbol{x} \in \mathcal{C}} \boldsymbol{c}^{\mathrm{t}} \boldsymbol{x}, \tag{3}$$

with the constraint set $\mathcal{C} = \{\boldsymbol{x} : \mathbf{M}\boldsymbol{x} \geq [1, \ldots, 1]^{\mathrm{t}}, \boldsymbol{x}^{\mathrm{t}}[1, \ldots, 1]^{\mathrm{t}} = N, x_\ell \in \{0, 1\}\}$, where x_ℓ is the ℓth element of $\boldsymbol{x}$. Since the value of x_ℓ is either 0 or 1 and there must be N 1's, an optimal solution of (3) can be considered as choosing N cost coefficients such that the cost is minimized while the constraint $\mathcal{C}$ is satisfied.

2.2 Dimensionality Reduction of EAP

Since the EAP has more feasible solutions than the 3DAP, it is not currently possible to solve (1) within a clinically acceptable time. In this paper, we use the same dimensionality reduction approach to the previous work [9], utilizing the following observation: the optimal solution has a near-zero cost when the c-arm pose error is low (it is zero when the pose is exactly known). This feature and Lemma 1 in [9] yield a following equivalent problem of reduced dimensionality (for proof, see [9, Sec. 2.2]):

$$\tilde{P}: \quad \min_{\tilde{\boldsymbol{x}} \in \tilde{\mathcal{C}}} \tilde{\boldsymbol{c}}^{\mathrm{t}} \tilde{\boldsymbol{x}}, \tag{4}$$

where $\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{c}} \in \mathbb{R}^K$ $(K \leq N')$ and $\tilde{\mathcal{C}} = \{\tilde{\boldsymbol{x}} : \tilde{\mathbf{M}}\tilde{\boldsymbol{x}} \geq [1, \ldots, 1]^{\mathrm{t}}, \tilde{\boldsymbol{x}}^{\mathrm{t}}[1, \ldots, 1]^{\mathrm{t}} = N, \tilde{x_\ell} \in \{0, 1\}\}$ with $\tilde{\mathbf{M}} = \mathbf{M}\mathbf{R}$ and where $\mathbf{R}$ is the *dimensionality reduction matrix* of size $N' \times K$ such that $\begin{bmatrix} x_{i1} & 0 & x_{i2} & 0 & \ldots & x_{iK} \end{bmatrix}^{\mathrm{t}} = \mathbf{R} \begin{bmatrix} \tilde{x}_1 & \tilde{x}_2 & \ldots & \tilde{x}_K \end{bmatrix}^{\mathrm{t}}$.

Given a solution $\tilde{\boldsymbol{x}}^*$ to the reduced problem $\tilde{P}$, the optimal solution to the original problem P is simply given by $\boldsymbol{x}^* = \mathbf{R}\tilde{\boldsymbol{x}}^*$.

2.3 Cost Coefficients and Seed Reconstruction

To compute $\mathcal{C}$, we need to compute the 3D intersection of the corresponding lines in space. Due to various errors (e.g., c-arm pose error, seed segmentation error, etc.), these straight lines never intersect, forcing us to compute a symbolic 3D intersection point. This point is typically defined as the global minimum of an error function. In this paper, we use *reconstruction accuracy* (RA) based on

the L_2 norm of Euclidean distance from the intersection point to the lines as a cost function. By using the RA cost, we can analytically compute a symbolic intersection point using a few summations followed by a 3×3 matrix inversion [3]. Once the EAP is solved, 3D locations of the N implanted seeds can be determined by the symbolic intersection points used to compute the N chosen RA costs in the solution (those that correspond to 1's in $\boldsymbol{x}$).

The dimensionality reduction approach described in Section 2.2 requires only the computation of K cost coefficients that are lower than a threshold. This implies that the exact value of most of the cost coefficients is not required. An efficient way to tell if a cost coefficient is higher than the dimensionality reduction threshold would allow us to skip its exact computation. This unnecessary cost computation can be avoided by utilizing the following Lemma.

Lemma 1. *Let I be the total number of 3D lines, and l_i and p_i be the unit direction vector of line i and a point on the line i, respectively. Every RA cost coefficient has the following lower bound:*

$$2I(I-1)RA \geq \sum_{i_1,i_2\in\{1,2,\ldots,I\},\ i_2>i_1} d(l_{i_1},l_{i_2})^2 \tag{5}$$

where RA is defined as $RA \triangleq 1/I\sum_{i=1}^{I}\|(P_I-p_i)\times l_i|^2$ and $d(l_{i_1},l_{i_2})$ is the Euclidean distance of line i_1 to line i_2 and with I images.

Proof. Due to the lack of space, the proof is not detailed.

Based on Lemma 1, we propose the following pruning algorithm.

Pruning algorithm for efficient computation of RA costs:

1: Compute every possible $d(l_{i_1},l_{i_2})^2$ for I images.
2: For the first i images, we have $\tilde{c}_{i_1,i_2,\ldots,i_i}=\sum_{i_1,i_2\in\{1,2,\ldots,i\},\ i_2>i_1} d(l_{i_1},l_{i_2})^2 = \tilde{c}_{i_1,\ldots,i_{i-1}}+\sum_{i_1=1}^{i-1} d(l_{i_1},l_i)^2$. Thus, $\tilde{c}_{i_1,i_2,\ldots,i_i}$ increases as the number of image i increases. When $\tilde{c}_{i_1,i_2,\ldots,i_i}>\eta$, the computation of $\tilde{c}_{i_1,i_2,\ldots,i_I}$ is not required and a large family of cost coefficients can be pruned. This favorable property allows for a recursive algorithm where images are virtually added one at a time and where a list of coefficients lower than η is updated.
3: Compute the RAs for indexes $(i_1,i_2,\ldots,i_I)$ remaining from step 2.

More RA cost coefficients are actually computed from the indexes of coefficients $\tilde{c}_{1,2,\ldots,I}$ lower than the dimensionality reduction threshold η, than are strictly required because (5) is only an inequality. The performance of the pruning algorithm directly depends on the ratio of the number of RA cost coefficients computed and the number of those actually lower than η. In practice, we observed that this ratio is in the range of 3 to 15, which is very compelling, due to its low cost.

2.4 Linear Programming

Based on the proposed dimensionality reduction, we solve our reduced integer program (4) using linear programming. We have implemented the linear program for the EAP using MATLAB command `linprog` followed by a test to confirm that its solution is binary (up to numerical errors).

3 Numerical Results

The algorithm was implemented using MATLAB 7.1 on a Pentium4 2.92 GHz PC with 3.8GB RAM. For simulations, we assumed that the camera parameters of the c-arm were known. For phantom and clinical datasets, we calibrated the c-arm prior to the procedure and computed the c-arm pose using a tracking fiducial that provides accuracy of 0.33° in rotation and 0.56 mm in translation [10].

3.1 Simulations

We performed simulation studies using synthetic projection images. We considered four cases with clinically realistic seed density of 2 and 2.5 seeds/cc and prostate size of 35 and 45 cc, resulting in 72, 84, 96, and 112 seeds. For each case, we generated three datasets. We generated six projection images on a 10° cone along the AP-axis in each dataset. In each image, there were 1.7% on average and up to 5.6% hidden seeds. We added random error to the pose, uniformly distributed on $[-h, h]$ (reported as h error). Rotation errors varied from 0° to 4°, with 1° steps and translation errors varied from 0 mm to 10 mm, with 2 mm steps. We exploited the fact that translation errors in depth are always significantly greater than those parallel to the x-ray image plane [3]. For each error type, 240 (4 cases × 3 datasets $\times C_3^6$) and 180 (4 cases × 3 datasets $\times C_4^6$) reconstructions were computed using three and four images, respectively. Shown in Fig. 1, our results imply that the EAP algorithm reliably finds the correct match and reconstruct the seeds with $> 95\%$ accuracy with up to 2° rotation error and 4 mm translation error even when only 3 images were used. When the c-arm pose is exactly known, the algorithm can localize the seeds with detection accuracy of $> 99\%$ and reconstruction error < 0.35 mm. We also conducted robustness tests on calibration errors varying from 0 to 10 mm and angular capture range

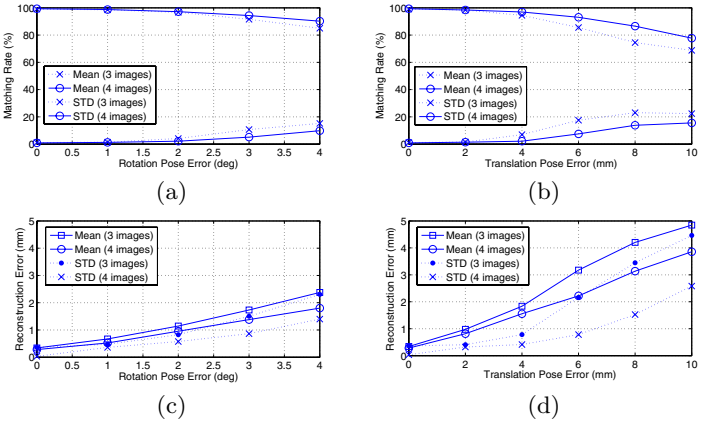


Fig. 1. Matching rates and reconstruction errors as functions of (a, c) rotation and (b, d) translation pose errors

Table 1. Phantom experiment results

Number of seeds	Mean match rate (%)	Mean ± STD reconstruction error (mm)	Mean computation time (s)
40	97.9	0.9 ± 0.7	1.1
55	100.0	0.6 ± 0.3	2.6
70	99.3	0.8 ± 0.3	3.9
85	97.6	1.1 ± 0.6	6.2
100	99.9	1.2 ± 0.5	10.8

varying from 5 to 25 degrees. Due to the limited space, we do not include the resulting plots, but our method is very robust to the calibration errors and small image acquisition angles achieving seed detection rate of $> 99\%$.

3.2 Phantom Experiments

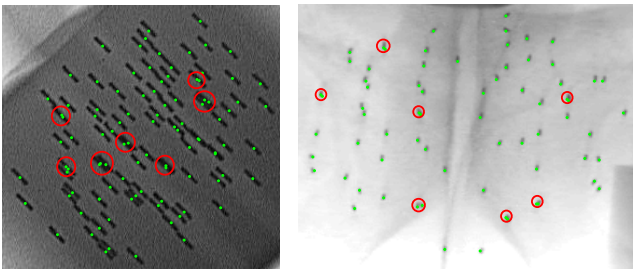
We evaluated the EAP algorithm on a precisely fabricated seed phantom assuring ground-truth. There were five datasets with 40, 55, 70, 85, and 100 implanted seeds (length of 4.9 mm and diameter of 0.8 mm, similar size to ^{125}I seeds) keeping seed density constant at 1.56 seeds/cc. For each dataset, we acquired six images within a 20° cone around the AP-axis using a *Philips Integris V3000*. On average, 5.5% and up to 22.5% of the seeds were hidden in each image. The EAP algorithm used 20 combinations of three images from the six available images in each dataset. It successfully localized the seeds with mean accuracy of $\geq 97.6\%$ and the mean reconstruction error of ≤ 1.2 mm within about 10 seconds or less depending on the number of seeds. Table 1 summarizes the results and Fig. 2(a) shows a phantom image example with re-projection of the detected seeds.

3.3 Clinical Experiments

We validated the EAP algorithm on six patient datasets. For each patient, we acquired two sets of images during the procedure, and one set of images at the end of the procedure using an *OEC 9800* fluoroscope. The image acquisition angle was about 20° around the AP-axis. Various number of ^{103}Pd seeds with length of 4.5 mm and radius of 0.8 mm were implanted (radio-opaque size of the x-ray marker is about three times smaller than the seed size). There were 0.9% on average and up to 7.8% hidden seeds in each image. Since we did not know the exact locations of the seeds, we visually assessed the correspondence between the projection of the estimated seeds and the actual seeds in the images and computed projection errors. The EAP algorithm almost perfectly localized the seeds using 3 images in all cases with mean projection error less than 1 mm within 10 seconds. We conservatively classified two seeds as mismatched, because the projection of them look matched in some images but look ambiguous in others. Thus, the true detection rate is between 98.1 % and 100 % - in either case, it is a clinically excellent performance. The reconstruction results are shown in Table 2 and Fig. 2(b) shows a fluoroscopy image example with re-projection of the detected seeds.

Table 2. Clinical experiment results

Patient ID	Number of seeds	Match rate (%)	Mean ± STD projection error (mm)	Computation time (s)
1	22	100	0.5 ± 0.3	1.1
	44	100	0.7 ± 0.4	2.2
	66	100	0.4 ± 0.2	4.8
2	39	100	0.3 ± 0.2	2.5
	82	100	0.3 ± 0.3	7.4
	84	100	0.9 ± 0.5	8.7
3	33	100	0.2 ± 0.1	1.4
	67	100	0.6 ± 0.5	3.5
	70	100	0.6 ± 0.6	4.7
4	35	100	0.9 ± 0.5	1.5
	68	100	0.2 ± 0.1	5.8
	77	98.7	0.5 ± 0.3	7.3
5	24	100	0.6 ± 0.6	1.0
	48	100	0.8 ± 0.6	2.4
	53	98.1	0.6 ± 0.5	2.7
6	33	100	0.3 ± 0.4	1.6
	61	100	0.1 ± 0.1	3.4
	66	100	0.1 ± 0.1	7.9



(a) (b)

Fig. 2. Fluoroscopy images with the re-projection of the estimated seed centroids (green dots). Red circles indicate overlapping seeds. (a) Phantom image with 100 seeds. (b) Patient image with 70 seeds.

4 Conclusion

We developed a computationally efficient and clinically feasible seed-matching algorithm for prostate brachytherapy seed localization. It automatically resolves the correspondence of seeds from multiple projection images with hidden seeds.

Simulation results imply that the EAP algorithm is robust to realistic c-arm pose errors with clinically acceptable accuracy. Especially in the phantom study, we used seeds that have similar radio-opaque size to ^{125}I seeds (about three times larger than that of ^{103}Pd seeds), thus creating more hidden seeds (up to 22.5%). On average, the EAP algorithm was able to correctly find the correspondence with matching rate of $\geq$ 97.6%. Even in the worst case where 22.5%, 7.5%, and 10.0% of the seeds are hidden in three images, it still achieved 92% correct matching rate. For clinical datasets, only ^{103}Pd seeds were used, thus having smaller number of hidden seeds than the phantom (^{125}I seeds).

The performance was almost perfect for all 18 datasets, and only two seeds were mismatched in two cases. Small projection errors (< 1.0 mm) imply the accuracy of our reconstruction. Note that a performance better than 97% detection rate and 2 mm reconstruction accuracy is considered to be clinically excellent.

The computation time was significantly reduced by adopting a new pruning method for efficient computation of the RA cost. Compared to the previously developed dimensionality-reduction-based algorithm [9] which solves the 3DAP in about 100 seconds, the EAP algorithm is more than 10 times faster and can solve the hidden seed problem within about 10 seconds. This is comparable to MARSHAL [3], one of the fastest seed-matching algorithms in the literature (and it solves an approximate formulation leading to a suboptimal solution).

Finally, we note that although the EAP algorithm is formulated for any number of images, simulation, phantom, and clinical experiment results show that three images are sufficient to achieve clinically adequate outcome in terms of accuracy, robustness, and computation time.

References

1. Altschuler, M.D., Kassaee, A.: Automated matching of corresponding seed images of three simulator radiographs to allow 3D triangulation of implanted seeds. Phys. Med. Biol. 42, 293–302 (1997)
2. Narayanan, S., Cho, P., Marks, R.: Fast cross-projection algorithm for reconstruction of seeds in prostate brachytherapy. Med. Phys. 29, 1572–1579 (2002)
3. Jain, A.K., Zhou, Y., Mustafa, T., Burdette, E.C., Chirikjian, G.S., Fichtinger, G.: Matching and reconstructoin of brachytherapy seeds using the hungarian algorithm (MARSHAL). Med. Phys. 32, 3475–3492 (2005)
4. Su, Y., Davis, B.J., Herman, M.G., Robb, R.A.: Prostate brachytherapy seed localization by analysis of multiple projections: Identifying and addressing the seed overlap problem. Med. Phys. 31, 1277–1287 (2004)
5. Narayanan, S., Cho, P.S., Marks II, R.J.: Three-dimensional seed reconstruction from an incomplete data set for prostate brachytherapy. Phys. Med. Biol. 49, 3483–3494 (2004)
6. Su, Y., Davis, B.J., Furutani, K.M., Herman, M.G., Robb, R.A.: Prostate brachytherapy seed reconstruction using adaptive grouping technique. Med. Phys. 34(7), 2975–2984 (2005)
7. Tutar, I.B., Managuli, R., Shamdasani, V., Cho, P.S., Pathak, S.D., Kim, Y.: Tomosynthesis-based localization of radioactive seeds in prostate brachytherapy. Med. Phys. 30, 101–109 (2003)
8. Lam, S.T., Cho, P.S., MarksII, R.J., Narayanan, S.: Three-dimensional seed reconstruction for prostate brachytherapy using hough trajectories. Phys. Med. Biol. 49(4), 557–569 (2004)
9. Labat, C., Jain, A.K., Fichtinger, G., Prince, J.L.: Toward optimal matching for 3D reconstruction of brachytherapy seeds. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 701–709. Springer, Heidelberg (2007)
10. Jain, A.K., Mustafa, T., Zhou, Y., Burdette, C., Chirikjian, G.S., Fichtinger, G.: FTRAC - a robust fluoroscope tracking fiducial. Med. Phys. 32, 3185–3198 (2005)

Prostate Biopsy Assistance System with Gland Deformation Estimation for Enhanced Precision

Michael Baumann[1,3,⋆], Pierre Mozer[2], Vincent Daanen[3], and Jocelyne Troccaz[1]

[1] Université J.Fourier, TIMC laboratory, Grenoble, France; CNRS, UMR 5525
[2] La Pitié-Salpêtrière hospital, urology dpt, 75651 Paris Cedex 13, France
[3] Koelis SAS, 5. av. du Grand Sablon, 38700 La Tronche, France
michael.baumann@imag.fr

Abstract. Computer-assisted prostate biopsies became a very active research area during the last years. Prostate tracking makes it possible to overcome several drawbacks of the current standard transrectal ultrasound (TRUS) biopsy procedure, namely the insufficient targeting accuracy which may lead to a biopsy distribution of poor quality, the very approximate knowledge about the actual location of the sampled tissues which makes it difficult to implement focal therapy strategies based on biopsy results, and finally the difficulty to precisely reach non-ultrasound (US) targets stemming from different modalities, statistical atlases or previous biopsy series. The prostate tracking systems presented so far are limited to rigid transformation tracking. However, the gland can get considerably deformed during the intervention because of US probe pressure and patient movements. We propose to use 3D US combined with image-based elastic registration to estimate these deformations. A fast elastic registration algorithm that copes with the frequently occurring US shadows is presented. A patient cohort study was performed, which yielded a statistically significant in-vivo accuracy of 0.83±0.54mm.

1 Introduction

Prostate biopsies are the only definitive way to confirm a prostate cancer hypothesis. The current clinical standard is to perform prostate biopsies under 2D TRUS control. The US probe is equipped with a needle guide for transrectal access of the prostate. The guide aligns the needle trajectory with the US image plane, which makes it possible to visualize the trajectory on the image for needle placement control. Unfortunately, in particular mid- and early-stage carcinoma are mostly isoechogenic, i.e. not visible in US images, which makes it necessary to sample the gland according to a systematic pattern. It is common to acquire 10 to 12 systematically distributed biopsies, the standard pattern taking roughly into account that most tumors (70%) develop in the peripheral zone of the gland.

⋆ Thanks to the Agence Nationale de la Recherche (TecSan project, France), the French Ministry of Health (PHRC program, France) and to Koelis S.A.S. (France) for funding.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 67–74, 2009.

The current standard biopsy procedure has several shortcomings: first, it is difficult for the clinician to reach systematic targets accurately because he has to move the probe continuously to place the needle; a constant visual reference is hence lacking. Second, performing a non-exhaustive systematic search for an invisible target implies that the target can be missed. Negative results leave the clinician in a dilemma when the cancer hypothesis cannot be discarded: his only option is to repeat the biopsy series. Furthermore, the location of the acquired samples with respect to the patient anatomy is only very approximately known after the intervention. Uncertainty about tumor location is the principal reason why prostate therapy is in general radical.

In order to address these issues, Baumann et al. [1] and Xu et al. [2] simultaneously proposed to acquire a US volume before the intervention and to use it as anatomical reference. The stream of US control images acquired during the intervention is then registered with the reference volume, which allows to project targets defined in the reference volume into the control images, and, conversely, the biopsy trajectory, known in control image space, into the reference volume. This technique makes it possible to improve biopsy distribution accuracy by showing the current trajectory in a fixed reference together with the trajectories of previously acquired biopsies, to aim targets defined in the reference volume during a planning phase, and to know the precise biopsy positions after the intervention. Non-US targets could originate from suspicious lesions in MR volumes that are then multi-modally registered with the US reference volume. It is also possible to derive targets from more sophisticated statistical atlases [3] or, in the case of repeated biopsies, they could consist of previously unsampled regions. After the intervention, the biopsy trajectories in the reference volume can be combined with the histological results and used for therapy planning.

Xu et al. [2] acquire a freehand 3D US volume and use 2D control images during the intervention. The 2D control images are tracked in operating room space with a magnetic sensor on the probe. In a second step, image-based registration is performed to compensate for small organ and patient movements. A similar approach was proposed by Bax et al. [4], who use an articulated arm for 2D US beam tracking. Bax does not compensate for patient and gland movements. However, pain-related pelvis movements are frequent, since the patient is not under total anesthesia. In that case, both methods risk to loose track of the gland because the US beam is tracked in operating room space and not in organ space, and a new reference volume has to be acquired. Baumann et al. address this draw-back by using 3D US to obtain richer control images during the intervention [1]. Instead of using a US beam tracking device to initialize local image-based registration, they propose a kinematic model of endorectal probe movements to compute anatomically plausible positions of the US beam with respect to the gland, which is unaffected by patient movements.

However, probe movements during needle placement continuously deform the gland. Deformations are strongest near the probe head and can reach 3 to 6 mm. They cannot be estimated with the presented systems. To address this issue, we extend the 3D US rigid registration approach presented in [1] by adding

a deformation estimation step to the registration pipeline. 3D US control images provide the information required to estimate the deformation with acceptable precision and accuracy. Inverse consistency and linear elasticity are used as deformation priors. A novel image distance measure capable of dealing with local intensity shifts, frequent in US images, is presented. The clinical accuracy of the presented algorithm is evaluated on a large number of patient data acquired during prostate biopsy sessions.

2 Method

Non-linear registration of 3D US image streams is currently the most promising approach to perform organ tracking with deformation estimation. The principal challenges of image-based tracking systems are robustness and computational efficiency. A technique to achieve both goals are coarse-to-fine registration strategies that successively increase the degrees of freedom (DOF) of the transformation space and the image resolution. In this paper, we add an elastic registration step to the 3-step coarse-to-fine rigid registration pipeline that we proposed in [1]. The resulting pipeline is illustrated in Fig. 1.

2.1 Framework for Non-linear Registration

Image-based deformation estimation can be formulated as an optimization process of a local distance measure. Let $I_1, I_2 : \mathbb{R}^3 \to \mathbb{R}$ be images, $\varphi : \mathbb{R}^3 \to \mathbb{R}^3$ the deformation function and the functional $\mathcal{D}[I_1, I_2; \varphi]$ a measure of the distance between I_1 and $I_2 \circ \varphi$. In contrast to parametric approaches that use basis functions to build the deformation function, we will follow a variational approach and define $\varphi(x) = x + u(x)$, where $u : \mathbb{R}^3 \to \mathbb{R}^3$ is assumed to be a diffeomorphism. The deformation could then be estimated by solving the optimization problem

$$\varphi^* = \arg\min_{\varphi} \left(\mathcal{E}[I_1, I_2; \varphi] \right), \tag{1}$$

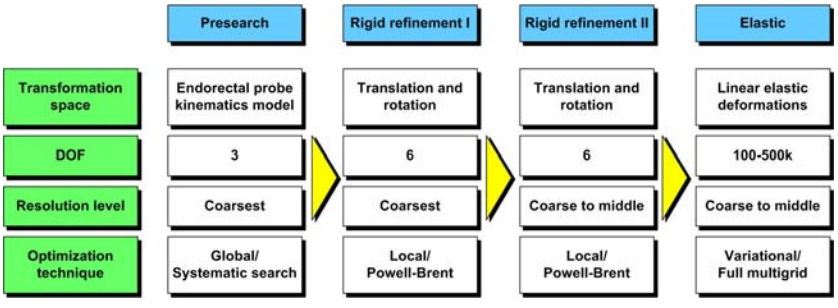


Fig. 1. Registration pipeline. The dimensionality of the transformation space and the image resolution are successively increased.

where the registration energy $\mathcal{E}$ simply corresponds to $\mathcal{D}$. Straightforward minimization of a distance measure yields, however, in general poor results due to countless local minima, in particular in presence of noise, partial object occlusion and other imperfections in the image data. Unfortunately, US is a particularly noisy modality, which makes 3D US based deformation estimation vulnerable to local misregistrations. This problem can be addressed by integration of a priori models of the expected deformation. This can be done implicitly by adding further energy terms to the objective function. In this work, inverse consistency and elastic regularization energies are added.

2.2 Inverse Consistency Constraints

In non-linear image registration, the forward estimation that minimizes $\mathcal{E}[I_1, I_2; \varphi]$ does in general not yield the inverse of the backward estimation that minimizes $\mathcal{E}[I_2, I_1; \psi]$, i.e. $\varphi \circ \psi \neq Id$ with $Id : \mathbb{R}^3 \to \mathbb{R}^3, x \mapsto x$. Introduction of Zhang's inverse consistency constraint [5]

$$\mathcal{I}[\psi; \varphi] = \int_\Omega ||\psi \circ \varphi - Id||^2_{\mathbb{R}^3} \; dx \tag{2}$$

as additional energy penalizes solutions that lead to inconsistent inverse transformations, where $\Omega \subset \mathbb{R}^3$ is the registration domain in image space. Estimation of the forward and the backward deformations is coupled by an alternating iterative optimization

$$\varphi^{k+1} = \arg\min_{\varphi} \left(\mathcal{E}[I_1, I_2; \varphi] + \mathcal{I}[\psi^k; \varphi] \right), \tag{3}$$

$$\psi^{k+1} = \arg\min_{\psi} \left(\mathcal{E}[I_2, I_1; \psi] + \mathcal{I}[\varphi^k; \psi] \right). \tag{4}$$

Concurrent estimation with mutual correction reduces the risk of local misregistrations.

2.3 Elastic Regularization

The deformation of the prostate caused by probe pressure is fully elastic, which justifies the introduction of the linearized elastic potential [6]

$$\mathcal{E}[\varphi] = \mathcal{E}[u + Id] = \int_\Omega \frac{\mu}{4} \sum_{j,k=1}^{3} \left(\partial_{x_j} u_k + \partial_{x_k} u_j \right)^2 + \frac{\lambda}{2} (\text{div } u)^2 \; dx \tag{5}$$

as additional energy, where λ and μ are the Lamé coefficients.

2.4 Image Distance Measure

The image distance measure is the driving energy of the optimization process. Experiments on patient data have shown that the sum of squared distances (SSD)

is a poor distance measure for deformation estimation on noisy US images. Local intensity changes are frequent due to changing US beam angles with respect to the tissues and probe pressure variations. The more robust Pearson correlation coefficient (CC) requires the evaluation of a large neighborhood of every voxel pair to yield statistically significant results, which is incompatible with deep multi-resolution approaches that operate on very coarse levels.

We hence prefer an intermediate correlation model that filters low-frequency intensity shifts, i.e. we assume that $I_1 \equiv I_2 \circ \hat{\varphi} + b$, where $\hat{\varphi}$ is the physical solution of the registration problem, and where $b : \mathbb{R}^3 \to \mathbb{R}^3$ models a local intensity shift. The shift is estimated by

$$b^\sigma[\varphi](x) = (I_1 - I_2 \circ \varphi) * \mathcal{G}_\sigma(x) \tag{6}$$

where $\mathcal{G} : \mathbb{R}^3 \to \mathbb{R}$ is a Gaussian with standard deviation σ. The image distance energy is then

$$\mathcal{D}[I_1, I_2; \varphi] = \int_\Omega (I_1(x) - I_2(\varphi(x)) - b^\sigma[\varphi](x))^2 \, dx. \tag{7}$$

The standard deviation σ controls the frequency range of the high-pass filter. If σ gets smaller, the cropped frequency range gets larger, and registration convergence rate decreases and may even stall if only high frequency noise like speckle is left. When used with a multi-resolution solver on a Gaussian pyramid (cf. next section), which implicitly performs a low-pass filtering of the intensity variations on coarse resolutions, this approach transforms to a band-pass filtering on varying frequency bands. In this configuration it is sufficient to chose relatively small standard deviations without risking registration inefficiencies.

2.5 Solver

Combination of the energy terms yields the alternating system

$$\varphi^* = \arg\min_\varphi \left(\mathcal{D}[I_1, I_2, \varphi] + \mathcal{E}[\varphi] + \mathcal{I}[\psi; \varphi]\right), \tag{8}$$

$$\psi^* = \arg\min_\psi \left(\mathcal{D}[I_2, I_1, \psi] + \mathcal{E}[\psi] + \mathcal{I}[\varphi; \psi]\right). \tag{9}$$

An iterative two-step minimization scheme is used to solve both objective functions. The Euler-Lagrange equations of Eqn. 8 and 9 are rewritten as a fixed point iteration

$$\frac{\varphi^{k+1} - \varphi^k}{\Delta t} = \mathcal{L}[\varphi^k] + f_\mathcal{D}[I_1, I_2; \varphi^k] + f_\mathcal{I}[\psi^k; \varphi^k], \tag{10}$$

$$\frac{\psi^{k+1} - \psi^k}{\Delta t} = \mathcal{L}[\psi^k] + f_\mathcal{D}[I_2, I_1; \varphi^k] + f_\mathcal{I}[\varphi^k; \psi^k], \tag{11}$$

where $t \in \mathbb{R}$ controls the discretization granularity, and with the elliptic partial differential operator

$$\mathcal{L}[\varphi] = \mathcal{L}[u + Id] = \mu \Delta u + (\lambda + \mu) \nabla \text{div } u, \tag{12}$$

which is obtained from the Gâteaux-derivative of $\mathcal{E}[\varphi]$ [6]. The Gâteaux derivatives of the energy term $\mathcal{D}$ at φ yields the force term

$$f_{\mathcal{D}}[I_1, I_2; \varphi] = (I_1 - I_2 \circ \varphi - b^{\sigma}[\varphi])(\nabla(I_2 - b^{\sigma})) \circ \varphi, \tag{13}$$

and for $\mathcal{I}$ we get

$$f_{\mathcal{I}}[\psi; \varphi] = (\psi \circ \varphi - Id)^T (\nabla \psi) \circ \varphi. \tag{14}$$

An iterative algorithm is used to estimate the displacement fields:

1: **while** not converged **do**
2: compute $f_{\mathcal{D}}[I_1, I_2; \varphi^k]$ and $f_{\mathcal{I}}[\psi^k; \varphi^k]$
3: compute $f_{\mathcal{D}}[I_2, I_1; \psi^k]$ and $f_{\mathcal{I}}[\varphi^k; \psi^k]$
4: solve Eqn. 10 for φ^{k+1}
5: solve Eqn. 11 for ψ^{k+1}
6: **end while**

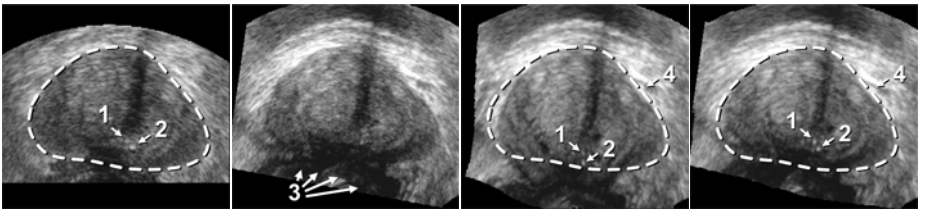
The forces are hence considered as constants for the resolution of the PDEs 10 and 11, and the forward and the backward estimation correct themselves mutually at each force update. The PDEs are solved using Red-Black Gauss-Seidel relaxation. Convergence is achieved if the difference of the L_2-norm of the total forces between two iterations is below a threshold for both the forward and the backward estimation (oscillatory states are detected). The algorithm is executed on various resolution levels of a Gaussian image pyramid [1] using the full multigrid strategy [7]. Note that the algorithm derives from the multigrid scheme by iterating until convergence on every grid level. This is necessary since the relaxation is performed on fractional forces. Fixed edges and bending side walls are used as border conditions [6]. The elasticity parameters are chosen such that Poisson's coefficient is zero, hence maximizing compressibility to allow compensation of local model inadequacies. Young's modulus is interpreted as a free variable in function of Poisson's coefficient and the PDE discretization Δt since it has no physical meaning in image registration. The forces are capped to a maximum length which makes it possible to control the maximum contributions per iteration to the displacement field via Δt. Limiting the contributions to less than 0.5 voxel side lengths ensures that the algorithm does not 'jump' over intensity barriers during optimization.

3 Experiments

The framework was validated on 278 registrations of 295 US volumes from 17 patients. The 17 reference images were acquired shortly before the intervention, and the tracking images were acquired after a biopsy shot. The clinical protocol was approved by the ethical committee of the XXX hospital, Town, Country, and all patients consented to participate to the study. The images were acquired with a GE Voluson and a RIC5-9 endorectal US probe. The algorithms were executed on a 4-core 2.6Ghz processor. In order to provide a reference gold standard for the evaluation of registration accuracy, experts manually segmented 467 point fiducials that were clearly identifiable on multiple images (e.g. calcifications and

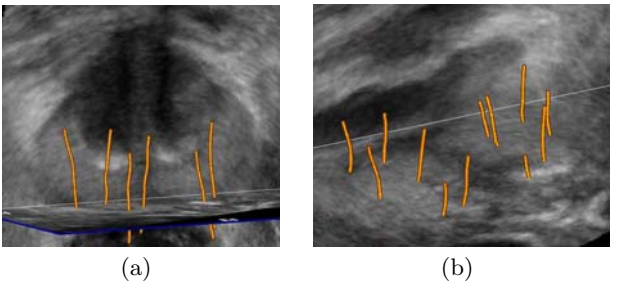
Table 1. Accuracy study

	mean distance	standard deviation	max distance	execution time (mean)
unregistered	13.76 mm	7.89 mm	51.61 mm	-
rigid	1.33 mm	0.85 mm	4.19 mm	2.1 s
elastic	0.83 mm	0.54 mm	4.14 mm	6.8 s



(a) (b) (c) (d)

Fig. 2. Standard SSD failure case study. Fig. (a) shows a prostate volume with calcifications [1,2]. Fig. (b) shows a second volume with a low contrast zone [3] after rigid registration with (a). Fig. (c) shows the 3D elastic registration with standard SSD and without inverse consistency; the whole prostate is dragged towards zone [3]; the tissue correspondences are worse than after rigid registration. Fig. (d) shows the 3D intensity shift filtered, inverse consistent elastic registration; the strong intensity differences between both volumes are correctly handled, the calcifications make appearance at the correct position (best viewed in PDF with zoom).



(a) (b)

Fig. 3. Tissue deformation corrected 3D biopsy maps in reference space

cysts). The distances between fiducial pairs were measured after registration to estimate the local accuracy. Note that the unavoidable segmentation error increases the measured error in average; this approach hence underestimates accuracy. Accuracy was computed for all registrations that were qualified as valid by experts after visual inspection, which represent 97,8% of the registrations. The results for both rigid and elastic registration are given in Tab. 1, and a visual illustration of the registration performance is given in Fig. 2. Fig. 3 shows 3D biopsy maps created with our biopsy tracking system.

4 Discussion and Conclusion

Deformation estimation achieves an overall accuracy of at least 0.83±0.54 mm on real patient data. This corresponds to an error reduction of 40% when compared to rigid 3D-3D registration. The average computation time of the registration was only 6.8s. We are confident that the algorithm can be accelerated to below 1s on the same machine with simple optimization and parallelization techniques, which is sufficient for assisted needle placement. With specialized standard hardware (GPUs), at least 5Hz should be feasible.

Biopsy tracking systems potentially add significant clinical value to prostate cancer diagnosis and therapy planning. Immediate advantages are the possibility to avoid resampling of already biopsied tissues when repeating a biopsy series, interventional quality control of the biopsy distribution (e.g. detection of unsampled areas) and computer-assisted guidance to non-systematic targets. The latter could for example be identified on MR/spectroMR images of the gland. Moreover, the improved knowledge about the biopsy and thus the cancer position could be used to implement focal therapy strategies for prostate cancer. 3D US based elastic tracking can provide the precision required for such therapeutic applications.

References

1. Baumann, M., Mozer, P., Daanen, V., Troccaz, J.: Towards 3D ultrasound image based soft tissue tracking: A transrectal ultrasound prostate image alignment system. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 26–33. Springer, Heidelberg (2007)
2. Xu, S., Krucker, J., Guion, P., Glossop, N., Neeman, Z., Choyke, P., Singh, A.K., Wood, B.J.: Closed-loop control in fused MR-TRUS image-guided prostate biopsy. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 128–135. Springer, Heidelberg (2007)
3. Shen, D., Lao, Z., Zeng, J., Zhang, W., Sesterhenn, I.A., Sun, L., Moul, J.W., Herskovits, E.H., Fichtinger, G., Davatzikos, C.: Optimization of biopsy strategy by a statistical atlas of prostate cancer distribution. Medical Image Analysis 8(2), 139–150 (2004)
4. Bax, J., Cool, D., Gardi, L., Knight, K., Smith, D., Montreuil, J., Sherebrin, S., Romagnoli, C., Fenster, A.: Mechanically assisted 3D ultrasound guided prostate biopsy system. Medical Physics 35(12), 5397–5410 (2008)
5. Zhang, Z., Jiang, Y., Tsui, H.: Consistent multi-modal non-rigid registration based on a variational approach. Pattern Recognition Letters, 715–725 (2006)
6. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, Oxford (2004)
7. Briggs, W., Henson, v.E., McCormick, S.F.: A multigrid tutorial, 2nd edn. Soc. for Industrial & Applied Math. (2008)

Prediction of the Repair Surface over Cartilage Defects: A Comparison of Three Methods in a Sheep Model

Manuela Kunz[1], Steven Devlin[1], Ren Hui Gong[1], Jiro Inoue[1], Stephen D. Waldman[1], Mark Hurtig[2], Purang Abolmaesumi[1], and James Stewart[1]

[1] School of Computing, Department of Materials and Mechanical Engineering, Department of Surgery, Queen's University, Kingston, ON, Canada
[2] Ontario Veterinary College, University of Guelph, Guelph, ON, Canada

Abstract. Defects in articular cartilage can be repaired through osteochondral transplantation (mosaic arthroplasty), where osteochondral plugs from non-weight-bearing areas of the joint are transferred to the defect site. Incongruity between the plug surface and the adjacent cartilage results in increased contact pressures and poorer outcomes. We compare three methods to predict the desired repair surface for use in computer-assisted mosaic arthroplasty: manual estimation, a cubic spline surface, and a statistical shape atlas of the knee. The cubic spline was found to most accurately match the pre-impact cartilage surface; the atlas was found to match least accurately.

1 Introduction

Articular cartilage injuries are one of the most common injuries seen in orthopaedic practice. In a review of 31,516 knee arthroscopies, it was found that cartilage defects appeared in 63% of these cases [1]. Damaged articular cartilage in weight-bearing areas of the knee has a large effect on the activity level and life style of the patient [2].

The limited self-healing potential of articular cartilage makes an operative management the common treatment for isolated cartilage defects. One well accepted surgical technique today is osteochondral autograft transplantation. During this procedure, cylindrical osteochondral plugs are retrieved from relatively non-weight-bearing areas of the knee and transplanted into the damaged regions.

For long-term success of this procedure, the transplanted plugs should reconstruct the curvature of the articular surface. Plugs that are placed too high will take a disproportionately large part of the joint load, which may lead to peak loading and abrasion of the plug surface and damage to the opposing articular surface. On the other hand, insufficient plug height results in inappropriate physiological pressure and may be disposed to late degradation [3].

Novel tools [4] and computer assisted procedures [5] have been used to achieve better plug harvesting and placement. But, in the presence of a cartilage defect

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 75–82, 2009.

(and the absence of the original cartilage surface), it can be challenging to achieve surface congruity due to lack of a good surface prediction mechanism, particularly in larger defects. Many effective methods have been proposed to segment cartilage from medical images (e.g. [6,7,8]), but we are not aware of any work to predict the original cartilage surface over a defect site.

The goal of this study was to compare three different artho-CT based methods for determining the desired articular repair surface for the use in computer-assisted cartilage repair surgeries. The three methods used manual estimation, a cubic spline, and a statistical shape atlas reconstruction. All methods were tested with an in-vivo sheep model study.

2 Material and Methods

Sixteen sheep knees – one from each sheep – were used in the study. For each knee, a CT scan was performed immediately after an intra-articular injection of a contrast agent. All arthro-CTs were obtained with a Light-Speed Plus CT (GE Healthcare, Waukesha, USA) in axial mode, with a slice thickness of 0.625 mm at 140 kpV. The sheep were positioned feet-first supine for the scan.

Three-dimensional surface models for bone and cartilage were manually built using the commercially available software package Amira (Visage Imaging, Inc., Carlsbad, CA, USA). An initial threshold segmentation was performed on the CT volume. This segmentation was manually refined using various editing functions. Isosurface models for bone and cartilage were created and stored.

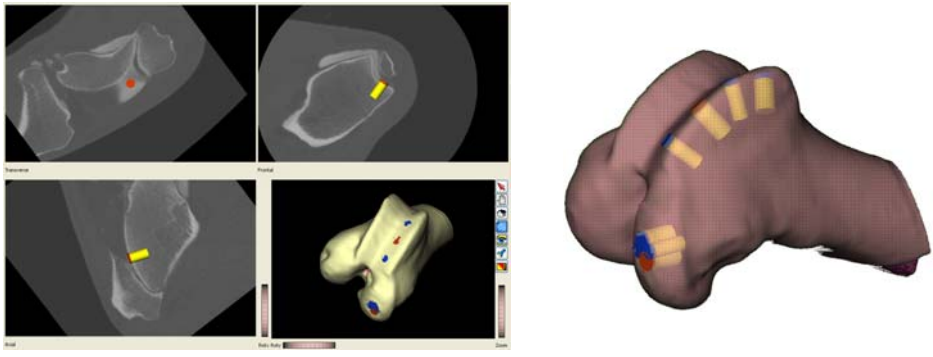
In a minimally-invasive surgical intervention, a cartilage defect on the medial condyle of each knee was induced with a calibrated impact. Three months after this cartilage damage, a second arthro-CT scan was obtained of each knee in-vivo using the same scanner and protocol as the pre-defect scans. Using the procedure described above, isosurface models for bone and cartilage were created. The three month defects varied in size between 40 mm^2 and 60 mm^2.

In each of the 16 post-defect CT datasets, the outline of the cartilage defect was manually defined and the desired articular surface over the defect area was predicted using each of three methods: **manual estimation**; **cubic spline surface**; and **statistical shape atlas**. We did not consider using the contralateral knee as a model because the morphology of the two knees is not alway identical, particularly when pathology is present.

The pre-defect and post-defect isosurfaces were co-registered using the iterated closest point (ICP) algorithm [9]. The distance between each predicted surface and the "gold standard" pre-defect isosurface was computed by selecting 50 points on the predicted surface and computing their root-mean-squared error to the pre-defect surface.

2.1 Method 1: Manual Estimation

Custom software was developed based on the freely available visualization library Coin3D (www.coin3d.org). For each knee, the post-defect CT volume and



(a) (b)

Fig. 1. (a) Manual planner for cartilage plug placement. The cartilage surface is visible below the light contrast agent in the CT slices. (b) The articular surface over the defect is predicted by the surfaces of the lower plugs on the condyle. (The upper four plugs are the corresponding donor sites.).

the 3D surface models for bone and cartilage were displayed to the operator. Our operator had several years of experience planning orthopaedic surgeries, including mosaic arthroplasty procedures. The operator identified the cartilage defect and created virtual cartilage plugs to reconstruct the articular surface.

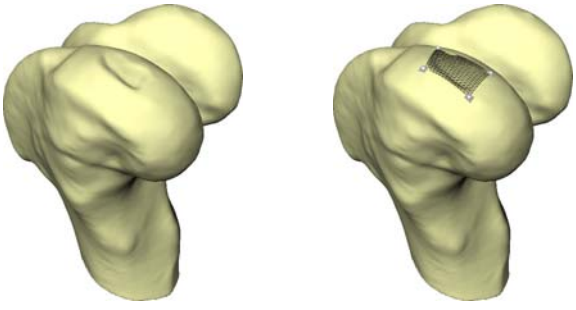
For each plug the radius, height, curvature, position and orientation could be modified by the operator. The operator manipulated these parameters to best fit the top of each cartilage plug to what the operator expected would be the pre-defect surface. Each plug could also be shown on the volumetric dataset in three orthographic planes, as shown in Figure 1(a). The operator did not see the pre-defect CT scans during this procedure. The surgeon verified that the planned plugs defined an acceptable articular surface over the defect.

Fifty evenly distributed points were automatically selected on the surfaces of the planned plugs. The resulting points defined the manually estimated cartilage surface and were compared with the gold standard surface.

2.2 Method 2: Cubic Spline Surface

A bicubic Hermite surface patch was applied over the defect site. This parametric patch, $Q(s,t)$, was defined by: four points on the surface; two non-orthogonal vectors, s and t, describing the direction between these points; the partial derivative with respect to s at each of those four points; the partial derivative with respect to t at each point; and the partial derivative with respect to *both* s and t at each point (this is called the "twist").

Another operator, who had no previously experience in planning of articular cartilage reconstruction, manually selected the four points $p_0, \ldots, p_3$ counterclockwise on the post-defect cartilage surface of each knee in a rough rectangle that completely surrounded the defect site, as shown in Figure 2. The vectors



(a) (b)

Fig. 2. (a) Cartilage defect before patch is placed. (b) Four corner points are selected by the operator to define a bicubic Hermite surface patch over the defect site.

s and t were defined as the direction between p_0 and p_1 and between p_0 and p_3. The partial derivatives at each point were computed using a central finite difference on the scanned cartilage surface, q:

$$\frac{\delta}{\delta s}q(s,t) \approx \frac{q(s+\Delta s,t) - q(s-\Delta s,t)}{2\Delta s} \tag{1}$$

with a Δs approximately 5% of the length of the corresponding edge of the rectangle. The twist vectors were set to zero, which results in a specific type of bicubic Hermite patch, called a Ferguson patch [10]. The surface patch was then constructed in the standard way (e.g. [11]):

$$Q(s,t) = \begin{bmatrix} F_1(s) \\ F_2(s) \\ F_3(s) \\ F_4(s) \end{bmatrix}^T \begin{bmatrix} q(0,0) & q(0,1) & q_t(0,0) & q_t(0,1) \\ q(1,0) & q(1,1) & q_t(1,0) & q_t(1,1) \\ q_s(0,0) & q_s(0,1) & q_{st}(0,0) & q_{st}(0,1) \\ q_s(1,0) & q_s(1,1) & q_{st}(1,0) & q_{st}(1,1) \end{bmatrix} \begin{bmatrix} F_1(t) \\ F_2(t) \\ F_3(t) \\ F_4(t) \end{bmatrix} \tag{2}$$

with $s,t \in [0,1]$ and $F_1(u) = 2u^3 - 3u^2 + 1$, $F_2(u) = -2u^3 + 3u^2$, $F_3(u) = u^3 - 2u^2 + u$, and $F_4(u) = u^3 - u^2$.

For each knee, the fifty points from the manually estimated surface of the same knee were projected onto the bicubic Hermite surface. These fifty projected points were roughly evenly distributed; they were compared to the gold standard surface.

2.3 Method 3: Statistical Shape Atlas

A statistical shape atlas was fit to the post-defect isosurface of the distal femur in order to predict the cartilage surface over the defect. The atlas used segmented geometric silhouettes from CT volumes containing the complete femur of 12 sheep cadaver legs (six left and six right). All 12 specimens came from sheep similar in age to the 16 sheep of the study group, but none of the 12 legs was from a sheep in the study group.

To avoid segmentation errors, all femurs were scanned in air. Each femur was dissected immediately after harvesting, keeping the cartilage on the knee intact. The cartilage was continuously hydrated using a phosphate buffered saline. The femur was scanned using a Light-Speed Plus CT (GE Healthcare, Waukesha, USA) in axial mode, with a slice thickness of 0.625 mm at 140 kpV. The CT images from left legs were mirrored.

Cartilage and bone were manually segmented from each of these training datasets and a binary CT volume (consisting of "inside" and "outside" voxels) was constructed for each knee. We used manual segmentation because the contrast agent is quite prominent in the image and varies substantially in location and shape, making intensity-based registration quite difficult.

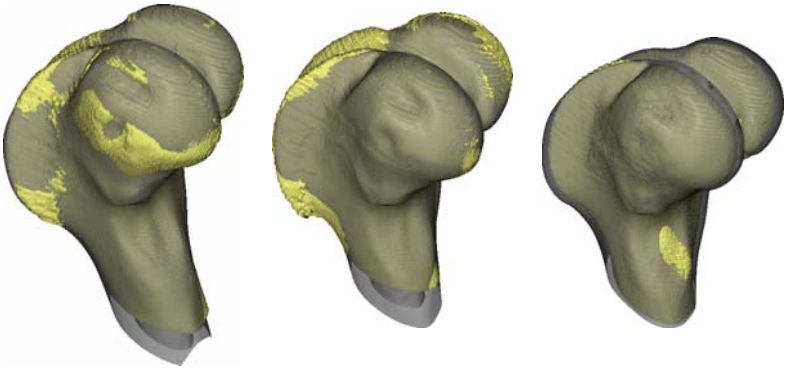
The atlas was built in two steps: data normalization and atlas construction. During the **data normalization step**, all training datasets were transformed into a common coordinate frame and decomposed into a mean shape plus a transformation that determined the variation from the mean shape. This transformation was modeled as the following sequential concatenation of transformations:

1. Rigid transformation for initial alignment: Each femur was reoriented such that its principal axes were aligned with the coordinate axes.
2. Anisotropic scaling for mean size calculation: The scaling factors for each femur were determined using the axis-aligned bounding boxes.
3. Rigid transformation and B-spline deformation for final alignment: A pairwise non-rigid registration was performed to generate the mean shape and calculate the final alignment.

For the **atlas construction step**, the statistical information of the transformations was captured. For each training dataset, the inverses of the transformations were computed, parameterized, and concatenated to form one column of a matrix, X. Since the inverse of a B-spline deformable transform is not analytically available, we approximated the inverse by registering the mean shape back to the training data. Finally, principal component analysis was performed on X and each training dataset, x_i, was projected into the eigen space, $a_i = diag(\delta_1 ... \delta_{12})([v_1 ... v_{12}]^T x_i)$, where $v_1, v_2, \ldots, v_{12}$ are the eigenvectors computed from XX^T, $\delta_1^2, \delta_2^2, \ldots, \delta_{12}^2$ are the corresponding variances along each eigenvector, and the a_i are the 12-dimensional atlas coefficients corresponding to x_i. The convex hull of all coefficients contained all valid shapes the atlas could generate, based on the set of training data.

To generate an instance of the atlas, a set of atlas coefficients was provided and the inverse process was performed to compute a sequence of transformations. These transformations were then applied in reversed order to the mean shape. To reduce computational costs, we used only the first eight eigen modes to generate an atlas instance, which accounted for about 85.5% of total variations in the training data (computed as $\sum_{i=1}^{8} \delta_i^2 / \sum_{i=1}^{12} \delta_i^2$)).

To **fit the atlas to a specific target knee**, the CT volume of the target knee (which had already been segmented) was converted into a binary volume of



(a) (b) (c)

Fig. 3. Samples of the atlas, as fitted to three different distal femurs. The atlas is shown as transparent gray over the yellow bone-and-cartilage surface. The defect is on the closer condyle of each femur. (a) and (b) show good fits, while (c) shows a poor fit in which the atlas surface is too far above the bone-and-cartilage surface (as seen on the silhouette).

"inside" and "outside" voxels. A rough initial alignment between the atlas mean shape and the target knee CT volume was made manually with a graphical user interface to bring the mean initial alignment RMS error to within 20 mm.

Then the CMA-ES optimization algorithm [12] was used to iteratively and simultaneously (a) refine the initial rigid alignment (6 parameters) and (b) determine the deformation of the atlas (8 parameters) using Mutual Information as the similarity metric applied to 5% (randomly selected) of the voxels of the target knee. The optimization algorithm used the extent of the target knee as the region of interest; the five percent of voxels were taken from this region and were drawn randomly with each iteration. Three sample results are shown in Figure 3.

For each knee, the fifty points from the manually estimated surface of the same knee were projected onto the atlas surface; the projected points, which were fairly evenly distributed, were compared to the gold standard surface.

2.4 Evaluation

For each of the 16 sheep knees, the cartilage surface was predicted using the three methods described above: manually, using a spline, and using an atlas. Each of these 48 predicted cartilage surfaces was compared to the corresponding true cartilage surface from the pre-defect arthro-CT scan of the same knee.

For each knee, the pre-defect and post-defect bone models were matched with a transformation computed using ICP; the same transformation was applied to the 50 points on each of the three predicted surfaces for that knee. The closest distance between each such point and the cartilage surface of the pre-defect scan was calculated and the RMS error of the 50 points computed.

Table 1. RMS errors, in mm, of the three methods on 16 knees

Method	mean RMS error 95% confidence interval	minimum RMS error	maximum RMS error
Spline	0.27 ± 0.09	0.10	0.54
Manual	0.30 ± 0.14	0.08	1.04
Atlas	1.16 ± 0.33	0.28	2.22

3 Results

Table 1 shows the RMS error between the surface obtained from each of the three methods and the pre-defect surface. The following conclusions were made by applying Student's t-test:

- The spline surface was more accurate than the atlas surface ($p < .0001$).
- The manual surface was more accurate than the atlas surface ($p < .0001$).
- The manual and spline surfaces had equivalent accuracy, considering an effect size of 0.25 mm RMS error to be insignificant ($p < .003$) (computed using the standard "two one-sided t-tests" method of testing for equivalence).

Also, we observed that the spline surface had a smaller variance in RMS error and a much smaller maximum RMS error than did the manual surface.

We did not measure the time taken to build each surface, but our operators reported the following **estimates** of time required to construct the predicted cartilage surface from the segmented pre-defect isosurface: 5 minutes for the spline surface; 20 minutes for the manual surface; and 110 minutes for the atlas surface. The manual time estimate included only the time to position the plugs on the defect site. The atlas time estimate used 80 iterations of the optimization procedure.

4 Discussion

We conclude that both the manual surface and the spline surface provide accurate reconstructions. The atlas surface, using our particular atlas, does not. The spline surface has several advantages over the manual surface: It is more consistently accurate, has smaller extreme variations, and is faster to use. Our results suggest that the cartilage surface can be well approximated by a cubic spline surface, similar to previous studies (e.g. [8]) that have used spline surfaces to enforce smoothness of the segmentation of an *existing* cartilage surface.

The poor performance of the atlas might be attributed to the construction or fitting of the atlas itself. More than 12 training knees would yield a better atlas. Restriction of the atlas to just the affected condyle might improve the matching. An atlas built from MR images has been shown [6] to provide excellent cartilage segmentation, and could be extended to predict the cartilage surface over defects. But the spline surface would likely remain superior due to its simplicity of implementation and fast surface construction time.

The study reported in this paper is part of a larger project on automated planning of computer-assisted mosaic arthroplasty. Any such automatic planner will need a computer representation of the predicted cartilage surface. It appears that a spline surface has the greatest advantages in such a system.

Acknowledgments. This research is supported by grant STPGP 336779-06 from the Natural Sciences and Engineering Research Council of Canada. The authors are grateful to Karen Lowerison, Nicole Kudo and Melissa Lui for their help with data collection and Emily Bishop, David Wright, John Li, Tamara Redwood and Jerome Grondin-Lazzazzera for their valuable help with segmenting the data. We wish to thank the reviewers for their helpful comments.

References

1. Curl, W., Krome, J., Gordon, E., Rushing, J., Smith, B., Poehling, G.: Cartilage injuries: A review of 31,516 knee arthroscopies. Arthroscopy: The Journal of Arthroscopic and Related Surgery 13(4), 456–460 (1997)
2. Jakob, R., Franz, T., Gautier, E., Mainil-Varlet, P.: Autologous osteochondral grafting in the knee: Indication, results, and reflections. Clinical Orthopaedics and Related Research 401, 170–184 (2002)
3. Koh, J., Wirsing, K., Lautenschlager, E., Zhang, L.Q.: The effect of graft height mismatch on contact pressure following osteochondral grafting: A biomechanical study. American Journal of Sports Medicine 32(2), 317–320 (2004)
4. Brzeczko, A., Goldberg, R., Taylor, R.H., Evans, P.: Smart alignment tool for knee mosaicplasty surgery. In: Medical Image Computing and Computer Assisted Intervention, pp. 599–605 (2001)
5. Hoser, C., Bichler, O., Bale, R., Rosenberger, R., Rieger, M., Kovacs, P., Lang, T., Fink, C.: A computer assisted surgical technique for retrograde autologous osteochondral grafting in talar osteochondritis dissecans (ocd): a cadaveric study. Knee Surgery, Sports Traumatology, Arthroscopy 12(1), 65–71 (2004)
6. Fripp, J., Crozier, S., Warfield, S., Ourselin, S.: Automatic segmentation of articular cartilage in magnetic resonance images of the knee. In: Medical Image Computing and Computer Assisted Intervention, pp. 186–194 (2007)
7. Dam, E., Folkesson, J., Pettersen, P., Christiansen, C.: Semi-automatic knee cartilage segmentation. In: SPIE Medical Imaging: Image Processing, pp. 1286–1294 (2006)
8. Cohen, A., McCarthy, D., Kwak, S., Legrand, P., et al.: Knee cartilage topography, thickness, and contact areas from mri: in-vitro calibration and in-vivo measurements. Osteoarthritis and Cartilage 7, 95–109 (1999)
9. Besl, P., McKay, N.: A method for registration of 3-d shapes. IEEE Transactions on Pattern Analysis and Machine Intelligence 14(2), 239–256 (1992)
10. Ferguson, J.: Multivariable curve interpolation. J. ACM 11(2), 221–228 (1964)
11. Salomon, D.: Curves and Surfaces for Computer Graphics. Springer, Heidelberg (2006)
12. Hansen, N.: The cma evolution strategy: A comparing review. In: Towards a new evolutionary computation. Advances in estimation of distribution algorithms, pp. 75–102. Springer, Heidelberg (2006)

A Coaxial Laser Endoscope with Arbitrary Spots in Endoscopic View for Fetal Surgery

Noriaki Yamanaka[1], Hiromasa Yamashita[1], Ken Masamune[1], Hongen Liao[2], Toshio Chiba[3], and Takeyoshi Dohi[1]

[1] Graduate School of Information Science and Technology, The University of Tokyo
[2] Graduate School of Engineering, The University of Tokyo
[3] Clinical Research and Development, National Center for Child Health and Development

Abstract. In this paper, we describe a rigid endoscope that transmits a laser beam coaxially to arbitrary points in the endoscopic view, mainly for treatment of twin-to-twin transfusion syndrome. The endoscope consists of a hotmirror for coaxial transmission of visible light and a Nd:YAG laser beam, and galvanometers for controlling the beam irradiation angle. We evaluated the transmission efficiency of the laser power, the spot size through the endoscope and accuracy in positioning the beam. The maximum laser transmission efficiency was 39% and the spot diameter was 2.2–3.2 mm at a distance of 10–20 mm. The positioning accuracy was mostly within 1.0 mm in the endoscopic view at the distance. The average laser power density on the spot was estimated to be 170–370 W/cm^2, and a chicken liver was successfully coagulated by changing the laser beam irradiation angle.

1 Introduction

Twin-to-twin transfusion syndrome (TTTS) occurs in about 15% of monochorionic twins, caused by imbalance in the net flow of blood between them across the placental communicating vessels [1,2]. The mortality rate of these twins exceeds 80%, and the risk of disability in the survivors is estimated to be 15–50% if untreated [3]. As a treatment for TTTS, fetoscopic laser photocoagulation (FLP) of the communicating vessels is widely accepted and the outcome showed that at least one survivor rate reaches 80 %.

During surgery, surgeons operate a fetoscope and a laser fiber in a sheath to navigate the laser guide-light to the communicating vessels. The laser is fired with a distance of 10 mm between the laser fiber tip and the vessels [5]. The operation requires skill, as it is necessary to avoid contact with the vessels or the placenta, which bleeds easily. Bleeding disrupts clear fetoscopic view, resulting in termination of the surgery. Surgeons have to stop bleeding as soon as possible. There exist other difficulties in this system. The distance is not precisely known from the fetoscopic image, and the vessels move as the mother breathes. In addition, the orientation between the fetoscopic image and the laser fiber is not easily identified.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 83–90, 2009.

In related studies, a composite-type optical fiberscope with a laser fiber coaxially coupled with optical image fibers has been developed. This fiberscope makes it easier to navigate the beam to the vessels because a laser beam is transmitted to the center of the view [6]. However, surgeons have to manipulate the fiberscope. The movement of the fiberscope must be small to avoid injuring the chorion and amniotic membranes, that possibly causing preterm rupture of the membranes. A safe treatment requires a technique for steering a laser beam on arbitrary targets without moving an endoscope. Furthermore, a thin fiberscope has a limited resolution and a narrow viewing angle and requires high illumination because light transmittance of the optical fiber is small. Surgeons are concerned that high illumination will affect premature fetal eyes.

Another related study stated a laser-pointing endoscope system that can acquire intraoperative 3D geometric information using two endoscopes, one projects a laser beam and the other captures the endoscopic image [7]. The projection point of the laser beam can be controlled using galvanometers. The laser source is a red diode which is easy to handle because its collimated beam is thin and the laser power is small. However, the system requires two endoscopes, which increases the invasion and the risk of complications. Furthermore, the low-powered laser doesn't suit photocoagulation. Therefore, development of an endoscope integrated with a high-powered laser irradiation function for arbitrary targets is important for treatment of TTTS.

We suggest a rigid endoscope that can transmit a laser beam to arbitrary points within the view without moving the endoscope itself. The rigid endoscope ensures a wide viewing angle and high transmittance of visible light.

2 Materials and Methods

2.1 Concept of Laser Endoscope

A rigid endoscope transmits a near-infrared laser beam for FLP, as shown in Fig. 1. Therefore, each lens of the endoscope has an anti-reflective (AR) coating for better transmission of visible and near-infrared light (VIS-NIR coating). For coaxial transmission of the laser beam, a hot mirror, which reflects near-infrared light and transmits visible light, is set between the endoscope and a camera. The beam and guide light from a laser fiber are reflected by galvano mirrors toward the hotmirror, which further directs this beam into the endoscope. The galvano mirrors control the direction of the beam at the entrance to the endoscope, then the beam is transmitted in a zigzag path around the optical axis of the endoscope and finally projected from the endoscope tip to some direction. The galvano mirrors and the camera are placed at optically symmetric points, which are equal to the eye point of the endoscope, with respect to the hot mirror to ensure a large range of the projection angle.

In clinical use, a surgeon selects a target point on the endoscopic image. the direction from the endoscope tip to the target is calculated, and the galvano mirrors rotate to appropriate angles to navigate the laser beam to the target. After confirming the guide-light spot on the image, the surgeon fires the laser.

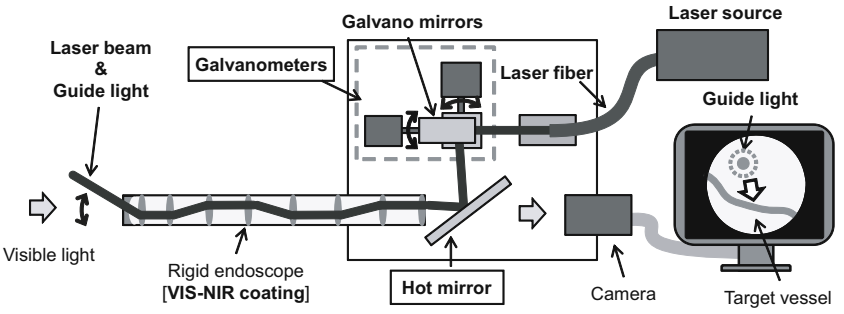


Fig. 1. Concept of the endoscope which provides an endoscopic image and transmits a laser beam coaxially to arbitrary target vessels in the endoscopic view

2.2 Design of the Laser Endoscope

To achieve a high laser power density in the spot, the laser spot should be small. A small spot can be achieved when the laser fiber has a small core and a small numerical aperture (NA) and the endoscope has object lenses with a small viewing angle and relay lenses with a large NA.

We used a Nd:YAG laser and a laser fiber (Fibertom 5100 / E-4070-B, Dornier MedTech) whose core diameter of 400 μm. The endoscope had relay lenses with an NA of about 0.12 and object lenses with a viewing angle of over 60 degrees. This design produced a spot as small as 2–3 mm in diameter at a distance of 10–20 mm from the endoscope tip with a changeable angle of the laser beam of about $\pm$15 degrees. The prototype endoscope consisted of concave or convex lenses with a diameter of 6 mm. The endoscope had an outer diameter of 7 mm, a viewing angle of about 70 degrees and an effective length of about 150 mm (Fig. 2). Handling was easy because its controlling unit was small and weighed as light as 450 g. The laser fiber tip was cooled by circulating water to protect it from damage by heat.

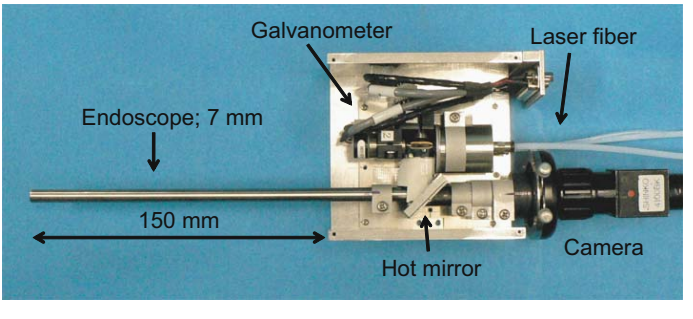


Fig. 2. Developed endoscope

2.3 Positioning the Laser Spot

For controlling the laser spot to target points indicated by the surgeon on the endoscopic image, it is necessary to calibrate the coordinates of the camera (or image) and the galvano mirrors. We acquired the relationship of their coordinates to the endoscope tip.

Camera Calibration. The relationship between a point $\boldsymbol{q}$ in the coordinates of the image and a point $\boldsymbol{p}$ in the coordinates of the endoscope tip is represented by a camera model with rotation R, translation $\boldsymbol{t}$, scale factors λ and α, and principal points u_o and v_o.

$$\lambda \begin{pmatrix} \boldsymbol{q} \\ 1 \end{pmatrix} = \begin{bmatrix} \alpha_u & 0 & u_o & 0 \\ 0 & \alpha_v & v_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & \boldsymbol{t} \\ O_{1\times 3} & 1 \end{bmatrix} \begin{pmatrix} \boldsymbol{p} \\ 1 \end{pmatrix} \tag{1}$$

We took fifty different postures of a checker pattern whose pitch was known, and calculated the coordinates and the distortion parameters by a flexible view technique [8]. The technique also provided the coordinates of the checker pattern to the camera. In addition, we used an optical tracking system (Micron Tracker 2, Claron Technology Inc.) to acquire the relationship between the camera and the endoscope tip. The optical tracker obtained the marker coordinates at the endoscope tip and the sheet where the checker pattern was written. The relationship between the checker pattern and the marker was known. Therefore, the relationship between the camera and the endoscope tip was calculated.

Galvano Mirror Calibration. The two galvano mirrors were located at different distances from the endoscope. We used a model, similar to the camera model, to calibrate each coordinate of the X- and Y-axes independently. The relationship between optical angle θ of the mirror and laser spot $\boldsymbol{p}$ in the coordinates of the endoscope tip is represented as below.

$$\lambda \begin{pmatrix} \tan\theta \\ 1 \end{pmatrix} = C_{in} \begin{bmatrix} R & \boldsymbol{t} \\ O_{1\times 3} & 1 \end{bmatrix} \begin{pmatrix} \boldsymbol{p} \\ 1 \end{pmatrix} \tag{2}$$

where

$$C_{in-X} = \begin{bmatrix} \alpha_X & 0 & u_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \; C_{in-Y} = \begin{bmatrix} 0 & \alpha_Y & v_o & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3}$$

A data-set of relations between the optical angles of the galvano mirrors and the position of laser spots was required to calculate the parameters. We ablated sheets fixed at some distances from the endoscope tip by rotating the galvano mirrors. We used the tracking system and acquired the coordinates of the markers on the sheets and the endoscope tip. The sheets were scanned and the relations of the ablated spots and the marker were recorded. Then, the positions of the spots were transformed into the coordinate of the endoscope tip. The transformation matrix from the target point to the angle of the mirrors was calculated by a technique similar to the direct linear transformation method after normalization of the data set [9,10].

3 Experiments and Results

We evaluated the transmission efficiency of laser power, the spot size, positioning accuracy of the spot and irradiated a chicken liver to test feasibility of the system.

3.1 Transmission Efficiency of Laser Power

We measured the power of the laser beam on the optical axis through the endoscope using a power meter (30(150)A-HE/NOVA, OPHIR) to evaluate the transmission efficiency. The laser power was set as 10–50 W at source. During measurement, the cooling water flowed at a speed of 6 ml/s. We didn't measure the power from the laser fiber directly at a source laser power of over 30 W because of the danger of breaking the fiber tip. Therefore, the transmission efficiency was calculated with the source power over 30 W.

The maximum efficiency was 39.1% at a source power of 10 W (Table 1). The efficiency decreased as the power increased. The maximum laser power from the endoscope was 16.1 W at a source power of 50 W.

Table 1. Transmission efficiency of the laser endoscope

Source	Laser fiber	Endoscope	Efficiency
10 W	10.9±0.1 W	4.3±0.2 W	39.1 %
20 W	20.5±0.1 W	7.9±0.1 W	38.6 %
30 W	-	11.1±0.1 W	36.9 %
40 W	-	13.6±0.1 W	34.1 %
50 W	-	16.1±0.1 W	32.2 %

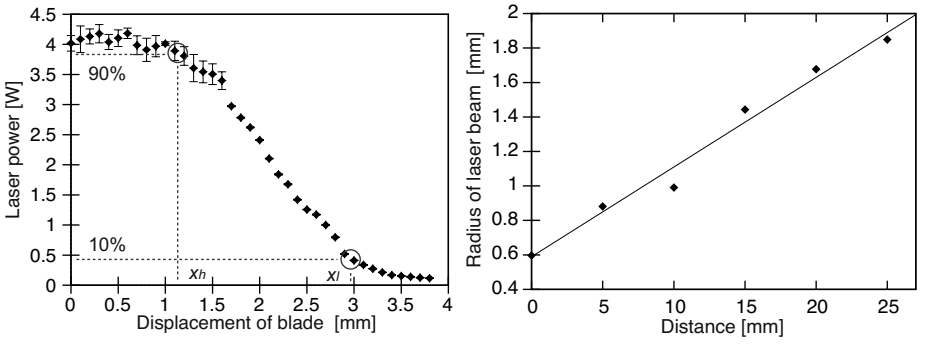


Fig. 3. Knife-edge method for measuring laser beam radius. Left: Laser power with blade displacement at a distance of 15 mm. Right: Laser beam radius from the endoscope tip.

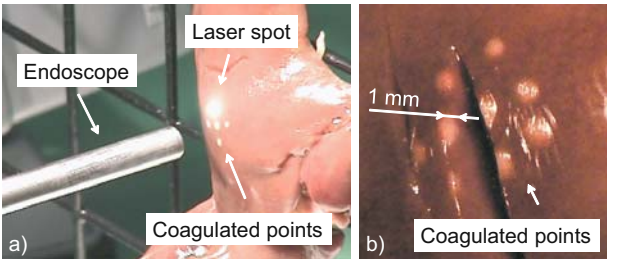


Fig. 4. Laser irradiation of a chicken liver. a) Changing irradiation angle; b) coagulated points on the liver.

3.2 Laser Spot Size

We evaluated the laser beam radius on the optical axis through the endoscope by the knife-edge method. A knife was set to cut the laser beam at a right angle and moved in 0.1 mm steps. The power meter behind the knife measured the portion of the laser not obstructed by the knife, and we captured the power shift from the displacement of the knife. The laser beam radius $d/2$ was calculated from displacement x_h, x_l at 90 and 10% of the maximum power (Fig. 3-left)[11].

$$d/2 = 1.561 \times |x_h - x_l| \ /2 \tag{4}$$

At a distance of 10–20 mm from the endoscope, the radius was about 1.1–1.6 mm, which equals 2.2–3.2 mm in diameter, as shown in the right side of Fig. 3.

3.3 Positioning Accuracy

We evaluated accuracy in positioning the laser spot. Sheets which have a grid with a pitch of 5 mm were placed at a distance of 10–20 mm from the endoscope tip. We pointed to the grid points on the endoscopic image and fired the laser. After irradiation, we scanned the sheet and measured the distance between the target and the actually irradiated points. The errors were 0.4±0.3 mm at a 10 mm distance, 0.5±0.4 mm at a 15 mm distance, and 0.5±0.6 mm at a 20 mm distance from the endoscope tip. The errors were mostly smaller than 1.0 mm in the view except on the upper side.

3.4 In Vitro Irradiation to Phantom

We irradiated a chicken liver with the laser to evaluate the laser endoscope for photocoagulation. A chicken liver is similar to a placenta in terms of the amount of blood. The liver was fixed at a distance of 10–15 mm from the endoscope and the laser source power was set as 50 W. We irradiated some points on the liver for 3 sec by changing the irradiation angle (Fig. 4-a). All the irradiated points were coagulated and the depths of coagulation were about 1 mm (Fig. 4-b).

4 Discussion

The maximum efficiency was nearly equal to the ideal value of 41% based on the transmission and reflection rates of the optical elements. Reduction in the transmission efficiency from the ideal value was due to dirt on the elements and power loss from the small galvano mirrors that could not use the full power of laser beam. At higher power, bubbles were generated around the laser fiber tip due to the cooling water, which diffused the beam. This resulted in loss of efficiency.

The laser spot diameter at a distance of 10–20 mm was 2.2–3.2 mm, which was suitable for coagulating placental vessels 1–2 mm in diameter. The distance is further than that of conventional treatment, which reduces the risk of contact with the placenta.

The laser endoscope achieved sufficiently high accuracy in positioning the laser spot because most of the errors were smaller than the diameter of the placental vessels in the target period for the surgery. In ordinary surgery, accuracy depends on the surgeon's skill, and there is risk in manipulating a fiberscope and a laser fiber. This risk is eliminated in the laser endoscope because the laser spot position can be controlled without moving the endoscope itself. The surgeon simply points to targets on the endoscopic image. This capability makes the surgery easier and safer.

Photocoagulation within a few seconds requires the power density of 100–1000 W/cm^2 [12,13]. The average power density of the spot was 170–370 W/cm^2 from 16.1 W at the source power of 50 W, that meets the minimum requirement. Furthermore, the endoscope successfully performed photocoagulation of a chicken liver.

For clinical use, we plan to miniaturize the endoscope by using smaller lens. A small laser spot can be achieved by an appropriate design. For higher transmittance, we optimize the selection of a laser and an AR coating and reduce the amount of lenses because the main cause of the transmission reduction is reflection of the laser beam on the surface of lenses.

5 Conclusion

We developed a prototype of a rigid endoscope that steers a laser beam on any point in the endoscopic view without moving the endoscope itself, mainly for treatment of TTTS. The maximum laser transmission efficiency was 39%. The spot diameter was 2.2–3.2 mm at a distance of 10–20 mm, which reduces the risk of contact with placenta. The positioning accuracy was mostly within 1.0 mm in the endoscopic view at the distance. The average laser power density at a spot was estimated to be 170–370 W/cm^2, and we successfully coagulated a chicken liver by changing the irradiation angle.

Acknowledgments. This study was partly supported by Japan Society for the Promotion of Science (JSPS, 21·9099) and Ministry of Education, Culture, Sports, Science and Technology (MEXT).

References

1. Sebire, N.J., Souka, A., Skentou, H., Geerts, L., Nicolaides, K.H.: Early prediction of severe twin-to-twin transfusion syndrome. Human Reproduction 15(9), 2008–2010 (2000)
2. Deprest, J., Jani, J., Lewi, L., Ochsenbein-Kölble, N., Cannie, M., Doné, E., Roubliova, X., Mieghem, T.V., Debeer, A., Debuck, F., Sbragia, L., Toelen, J., Devlieger, R., Lewi, P., Velde, M.V.: Fetoscopic surgery: Encouraged by clinical experience and boosted by instrument innovation. Seminars in Fetal and Neonatal Medicine 11, 398–412 (2006)
3. Yamamoto, M., Murr, L.E., Robyr, R., Leleu, F., Takahashi, Y., Ville, Y.: Incidence and impact of perioperative complications in 175 fetoscopy-guided laser coagulations of chorionic plate anastomoses in fetofetal transfusion syndrome before 26 weeks of gestation. Am. J. Obste. Gynecol. 193, 1110–1116 (2005)
4. Rossi, A.C., D'Addario, V.: Laser therapy and serial amnioreduction as treatment for twin-twin transfusion syndrome: a metaanalysis and review of literature. Am. J. Obste. Gynecol. 198, 147–152 (2008)
5. Ville, Y., Hecher, K., Ogg, D., Warren, R., Nicolaides, K.: Successful outcome after Nd:YAG laser separation of chorioangiopagus-twin under sonoendoscopic control. Ultrasound Obstet. Gynecol. 2, 429–431 (1992)
6. Oka, K., Naganawa, A., Yamashita, H., Nakamura, T., Chiba, T.: Composite-Type Optical Fiberscope for Laser Surgery for Twin-to-Twin Transfusion Syndrome. In: Dohi, T., Sakuma, I., Liao, H. (eds.) MIAR 2008. LNCS, vol. 5128, pp. 251–259. Springer, Heidelberg (2008)
7. Hayashibe, M., Nakamura, Y.: Laser-Pointing Endoscope System for Intra-Operative 3D Geometric Registration. In: Proceedings of the 2001 IEEE International Conference on Robotics and Automation, vol. 2, pp. 1543–1548 (2001)
8. Zhang, Z.: A Flexible New Technique for Camera Calibration. IEEE Transaction on Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)
9. Abdel-Aziz, Y.I., Karara, H.M.: Direct linear transformation from comparator coordinates into object space coordinates. In: Proceedings of the ASP/UI Symposium on Close-Range Photogrammetry, pp. 1–18 (1971)
10. Marzan, G.T., Karara, H.M.: A computer program for direct linear transformation solution of the collinearity condition, and some applications of it. In: Proceedings of the Symposium on Close-Range Photogrammetric Systems, pp. 420–476 (1975)
11. Siegman, A.E., Sasnett, M.W., Johnston Jr., T.F.: Choice of Clip Levels for Beam Width Measurements Using Knife-Edge Techniques. IEEE Journal of Quantum Electronics 27(4), 1098–1104 (1991)
12. Boulnois, J.-L.: Photophysical Processes in Recent Medical Laser Developments: a Review. Lasers in Medical Science 1, 47–66 (1986)
13. Dörschel, K., Brodzinski, T.: Proposal for dosimetry of non-ionizing radiation. Lasers in Medical Science 4(1), 329–340 (1989)

Toward Video-Based Navigation for Endoscopic Endonasal Skull Base Surgery

Daniel Mirota[1], Hanzi Wang[2], Russell H. Taylor[1], Masaru Ishii[3], and Gregory D. Hager[1]

[1] Department of Computer Science, The Johns Hopkins University, Baltimore, MD, USA
[2] School of Computer Science, The University of Adelaide, SA, Australia
[3] Department of Otolaryngology, Johns Hopkins Medical Institutions, Baltimore, MD, USA
dan@cs.jhu.edu

Abstract. Endoscopic endonasal skull base surgery (ESBS) requires high accuracy to ensure safe navigation of the critical anatomy at the anterior skull base. Current navigation systems provide approximately $2mm$ accuracy. This level of registration error is due in part from the indirect nature of tracking used. We propose a method to directly track the position of the endoscope using video data. Our method first reconstructs image feature points from video in 3D, and then registers the reconstructed point cloud to pre-operative data (e.g. CT/MRI). After the initial registration, the system tracks image features and maintains the 2D-3D correspondence of image features and 3D locations. These data are then used to update the current camera pose. We present registration results within 1mm, which matches the accuracy of our validation framework.

1 Introduction

Endoscopic endonasal skull base surgery (ESBS) has gained much interest recently over traditional open surgical approaches as treatment for cancers of the sinus and pituitary gland. Pituitary lesions, though generally benign, are the most common brain tumor. These common pituitary lesions, as well as cancers of the nasal cavity, brain cancers surrounding the nose, and cancers involving both the nose and brain are all treated with ESBS. Treatment with traditional surgical approaches to the skull base are associated with significant morbidities because healthy cranial nerves are sometimes damaged during surgery. Unlike traditional approaches, ESBS is less invasive and is shown to reduce operative time as well as decrease the length of hospital stay [1].

ESBS and traditional approaches are best contrasted with a clinical example. Figure 1 shows a coronal MRI scan of a patient with a pituitary macroadenoma. The central location of this tumor makes it difficult to approach using traditional means. The tumor is flanked by the carotid arteries and the optic chiasm, and the left optic nerve is clearly compressed by tumor. This tumor was removed using an endoscopic endonasal approach. The endoscopic image (insert) was taken just after the tumor was removed. Notice that the optic nerve (ON) has dropped significantly in height–almost to the level of the carotid artery (CA). Manipulating such high-value structures in cases like this are the reason that ESBS remains a painstaking procedure that requires precise knowledge of patient anatomy. Thus, surgical navigation is key for success, especially to aid junior

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 91–99, 2009.

surgeons and for complex cases [1], as it provides the surgeon a means to both maintain orientation and monitor progress.

In current practice, the surgeon uses a pointer tool to interact with the navigation system. The system tracks rigid-bodies attached to the tool and the patient. During preparation for surgery the rigid-body attached to the patient is registered to fiducial markers on the patient. The rigid-body defines the patient's head in the navigation system. The navigation system in turn calculates a rigid-body transformation between the patient's head and the tool to display the corresponding location in CT. A drawback of the current procedure is that each rigid-body transformation measurement contributes to localization error. In fact, localization error is typically quoted as $2mm$ with a good registration, and can be far larger with a poor one [2]. Errors of this magnitude could lead to surgical error resulting in high morbidity or in mortality.

To improve current navigation systems, we propose a new system that utilizes endoscopic video data for navigation. While endoscopic video presents many challenges including reflection, specularity, and low texture, our system robustly handles these challenges and creates a 3D reconstruction from video. Then the system registers the 3D reconstruction to a pre-operative CT scan. After the initial registration the system tracks the camera location by matching image features and performing robust 2D-3D pose estimation. Instead of relying on the long rigid-body transformation chain that current navigation systems use, video-CT registration employs a more direct accurate localization of the camera relative to the patient. We show tracking results within millimeter mean error of an Optotrak (Northern Digital Corp. Waterloo, Canada) measured camera motion used for comparison.

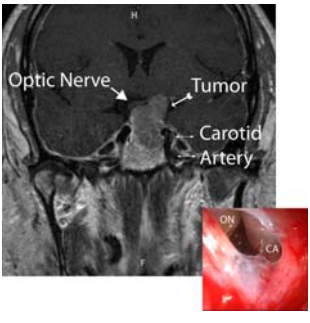


Fig. 1. An example coronal MRI of a patient and endoscopic image (insert) of how clear identification of the carotid artery and optic nerve is crucial

Tracking the location of a camera relative to CT has been studied in other areas of image-enhanced surgical navigation. In 2002, Shahidi et al. [3] presented a system for endoscope calibration and image-enhanced endoscopy. They achieved millimeter accuracy for a system using passive optical markers for tracking. More recently, Lapeer et al. [4] evaluated another similar system again using passive optical markers for tracking and reported that submillimeter accuracy still remains elusive. Video registration has been previously applied to bronchoscopy [5] where normalized mutual information was used to register to CT. In [6], visual tracking and registration was demonstrated on a skull phantom.

2 Method

There are five major components in our system. Figure 2 shows an overview of the system with the five components highlighted in blue. First, the system extracts SIFT features [7] from the video data. Next, the motion between the images is estimated, after which feature points are reconstructed. At the beginning of the video sequence, reconstructed points from the first pair of images are registered to a segmented surface in the

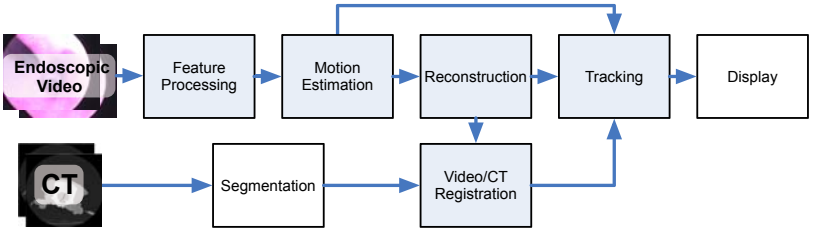


Fig. 2. System Overview

CT image. This initial registration is used to initialize a registration tracking algorithm that makes use of feature matches in subsequent frames. Each of these components is detailed in the following subsections.

2.1 Feature Processing, Motion Estimation and Reconstruction

Before reconstruction or registration, video data are processed to extract image features using the Matlab implementation [8] of SIFT features [7]. We tested both the SIFT feature matching technique suggested by Lowe and SVD SIFT matching technique [9]. While SVD SIFT provides more matches, the matches are not of the same accuracy as Lowe's. However, the larger number of matches increases the number of points in the reconstruction. Figure 3 shows the difference in the two matching methods.

After the image features are detected and matched, the motion is estimated using the robust technique of [10]. To reconstruct points in two or more images the rigid-body transformation between the images must be computed. One 3D point imaged in two images separated by a rotation, R, and translation, $\mathbf{t}$, forms the epipolar constraint. Given a calibrated point in image 1, p_1, and a calibrated point in image 2, p_2, the epipolar constraint is written as

$$p_2 E p_1 = 0 \tag{1}$$

Where $E = sk(\mathbf{t})R$ is known as the Essential Matrix and $sk(\mathbf{t})$ is the skew-symmetric matrix of $\mathbf{t}$. Here, we solve (1) using the robust E-matrix estimator from [10] that uses the five-point algorithm [11]. Motion is estimated by finding the E-matrix that best matches a set of inliers.

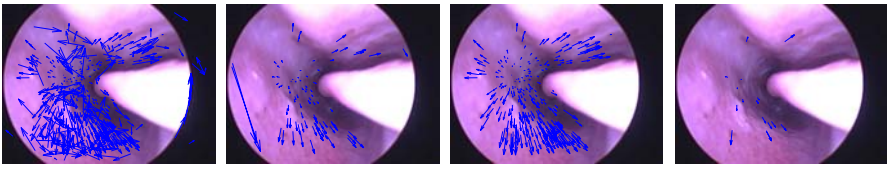


Fig. 3. From left to right: SVD SIFT matches, SIFT matches, inliers of SVD SIFT matches and inliers of SIFT matches

Once the motion estimation is completed, the motion is used to reconstruct the 3D structure. The first pair of images is reconstructed with triangulation and subsequent pairs of images are reconstructed from the tracked image features. The image features are tracked using SVD SIFT feature matches.

2.2 Registration and Tracking

The reconstructed 3D point cloud is registered to a surface segmented from a CT image of the same patient. The surface is segmented by applying a threshold at the air/tissue boundary and using marching cubes to create the isosurface. We applied a registration algorithm described in [12] that is derived from Trimmed ICP (TrICP) [13] and extends TrICP with scale [14]. The registration algorithm needs to estimate scale because the true scale of the 3D world is lost in the epipolar constraint (1).

After the initial 3D-3D registration, the 2D-3D correspondence between image features and the 3D surface of the CT is established. Now, a more efficient 2D-3D pose estimator can be used to update the camera pose. Here we combine the robust sampling method of [10] with the pose algorithm of [15] to create a robust 2D-3D pose estimation method. In Algorithm 1, we present an overview of the complete system. $R_{init}, \mathbf{t}_{init}, s_{init}$ are the initial rotation, translation and scale respectively. $points$ is the initial sparse 3D reconstruction. F is a set of images from the video. $mesh$ is the surface mesh segmented from the CT data.

Algorithm 1. $(R, \mathbf{t})$ = Tracking$(R_{init}, \mathbf{t}_{init}, s_{init}, points, F, mesh)$

$R \leftarrow R_{init}, \quad \mathbf{t} \leftarrow \mathbf{t}_{init}, \quad currentPoints \leftarrow s_{init} points$
for all $f_1, f_2 \in F$ where f_1, f_2 are 3 frames apart **do**
  $(\hat{f}_1, \hat{f}_2)$ =undistort(f_1, f_2)
  (sift$_1$, sift$_2$) = detect SIFT feature$(\hat{f}_1, \hat{f}_2)$
  $matches$ = SVDSIFT match(sift$_1$, sift$_2$)
  $(E, inliers)$ = robustMotionEstimator($matches$)
  $currentPoints$ = tracker($matches$,$inliers$,$currentPoints$)
  $(\hat{R}, \hat{\mathbf{t}})$ = robustPoseEstimator($currentPoints, matches, R, \mathbf{t}$)
  $reprojectedInliersPoints$ = reprojectPoints($currentPoints, \hat{R}, \hat{\mathbf{t}}$)
  $(R, \mathbf{t})$ = robustPoseEstimator($reprojectedInliersPoints, matches, \hat{R}, \hat{\mathbf{t}}$)
  $refinedPoints$ = refinePoints($reprojectedInliersPoints$);
  for all previous $R_i, \mathbf{t}_i$ **do**
    $(R_i, \mathbf{t}_i)$ = robustPoseEstimator($refinedPoints_i, matches_i, R_i, \mathbf{t}_i$)
  end for
end for

The system first undistorts the images, then SIFT features are detected and matched. After finding the set of matched image features, these image feature points are used to estimate the motion of the frame pair. The inliers of the motion estimation are then tracked using the SIFT feature tracking method from [10]. Once the new camera pose is estimated, we applied one of the following two methods to refine both the 3D points

and pose estimates. Method I, applies all of the previously seen image features and projection matrices to determine the refined 3D point by setting up the following null space problem. First, given a simple projection model $\begin{bmatrix} u_i & v_i & 1 \end{bmatrix}^T = \begin{bmatrix} R_i & \mathbf{t}_i \end{bmatrix} \begin{bmatrix} x_i & y_i & z_i & 1 \end{bmatrix}^T$.

Let $R_i = \begin{bmatrix} r_{1,i}^T & r_{2,i}^T & r_{3,i}^T \end{bmatrix}^T$ and $\mathbf{t}_i = [t_{1,i}\ t_{2,i}\ t_{3,i}]^T$. Then the 3D point which was created by all of the imaged points is the null space solution of

$$0 = \begin{bmatrix} r_{1,1}^T\ t_{1,1} - u_1\, r_{3,1}^T\ t_{3,1} \\ r_{2,1}^T\ t_{2,1} - v_1\, r_{3,1}^T\ t_{3,1} \\ \vdots \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{bmatrix} \tag{2}$$

Method II, uses the initially registered camera position to raycast image feature points of the same 3D point onto the CT scan and calculates the mean of the projected points. We compute each ray as follows. First given the camera's intrinsic parameters, K, rotation, R, and translation, $\mathbf{t}$. We project a point, p, by applying the intrinsic parameters, $p_n = K^{-1}p$. Then the ray is $p_r = (Rp_n + \mathbf{t}) - \mathbf{t}$. The refined points are then used with the robust 2D-3D pose estimator to compute the new camera pose.

3 Experiments

3.1 Data Collection

We collected endoscopic ex-vivo porcine sinus video data. The video was captured at 640x480 using framegrabber attached to a Storz Telecam, 202212113U NTSC with a zero-degree rigid monocular endoscope. Optotrak rigid-bodies were attached to the endoscope and porcine specimen. The Optotrak and video data were simultaneously recorded. The Optotrak motion data were used as the ground truth to compare with the estimated endoscopic motion. Before the data collection, images of a checkerboard calibration grid were also recorded using the endoscope. We performed an offline camera calibration of the endoscope using the Matlab Camera Calibration Toolkit [16]. After the camera calibration, the hand-eye calibration between the Optotrak rigid-body and the camera was estimated by solving the AX=XB problem using the method from [17]. The CT data used had 0.5x0.5x$1.25mm$ voxels.

Our data collection had three major sources of error: one, the camera calibration; two, the Optotrak to camera calibration; and three the Optotrak to CT registration. The camera was calibrated within a pixel error of $[0.38, 0.37]$. The aforementioned Optotrak/camera configuration had an estimated position uncertainty in millimeters of the camera center as measured by the Optotrak of $[1.1, 0.4, 0.1]$. Finally, the Optotrak to CT registration was generally within $.5mm$ RMS point distance error. Each of these contributes to an overall location uncertainty of approximately $1.5mm$ in the absolute position of the endoscope optical center, and approximately $1.1mm$ relative position accuracy.

3.2 Results

We tested our algorithm on 14 randomly selected segments of video from a porcine sinus. Each segment was from a different region of the sinus from anterior to posterior shown

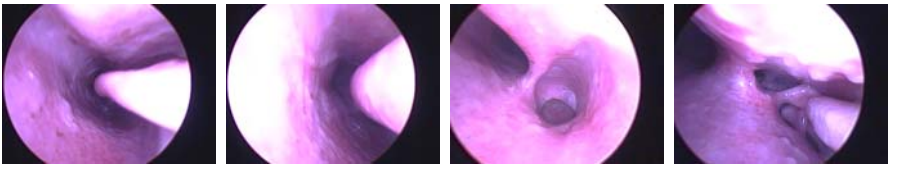


Fig. 4. From left to right the first image of video segments 1 through 4

in figure 4. Figure 4 also shows the challenges of these images including specularities, highlights, motion blurring, lack of texture, and over saturation.

In figure 5 we present both our estimated and Optotrak measured trajectories as well as error along the trajectories for video segment 1. The error is computed as follows. Each image has both an estimated pose and pose measured by the Optotrak. For example $R_1, \mathbf{t}_1$ and $R_{O1}, \mathbf{t}_{O1}$ where $_O$ denotes an Optotrak measurement. The error is then given by $F_{error} = \begin{bmatrix} R_{O1} & \mathbf{t}_{O1} \\ \mathbf{0}^T & 1 \end{bmatrix}^{-1} \begin{bmatrix} R_1 & \mathbf{t}_1 \\ \mathbf{0}^T & 1 \end{bmatrix}$. The translation error is the l_2-norm of the translation component of F_{error}. The rotation error is the absolute sum of the Euler angles of the rotation component of F_{error}. The results are aligned to the Optotrak pose.

Of the 14 segments, one failed to maintain at least five points and three diverged. The three that diverged did not have sufficient 3D structure in the scene for initial registration. The overall accuracy of the remaining 10 achieved by the proposed algorithm is within approximately 1 millimeter of the Optotrak measurement of relative motion. That is, our results are within the error envelope of our measurement system itself. The quantitative tracking results are shown in table 1. We compare our method versus the Optotrak visually in figure 6 which reveals that the Optotrak (the right of each pair) does not align as well as the visually registered image.

Table 1. Mean pose error and standard deviation along of the trajectories

	Trans. Error (mm)		Rot. Error (deg)	
	Method 1	Method 2	Method 1	Method 2
Segment 1	0.83 (0.38)	1.08 (0.52)	1.42 (0.94)	1.39 (1.08)
Segment 2	0.90 (0.40)	0.94 (0.74)	1.07 (0.86)	1.80 (1.90)
Segment 3	0.94 (0.76)	0.51 (0.40)	0.66 (0.56)	1.46 (1.34)
Segment 4	1.24 (0.83)	0.96 (0.52)	1.79 (1.00)	0.91 (0.84)
Segment 5	0.64 (0.34)	0.58 (0.29)	1.14 (0.98)	0.65 (0.39)
Segment 6	0.41 (0.21)	0.75 (0.48)	1.93 (1.02)	1.49 (0.96)
Segment 7	1.32 (0.97)	2.5 (2.21)	3.68 (1.54)	6.29 (4.34)
Segment 8	0.40 (0.24)	0.33 (0.22)	1.79 (1.09)	1.54 (1.43)
Segment 9	1.03 (0.72)	0.43 (0.29)	0.59 (0.24)	0.66 (0.46)
Segment 10	0.56 (0.31)	0.55 (0.31)	2.94 (2.15)	0.90 (0.54)
Mean Error	0.83	0.86	1.70	1.71

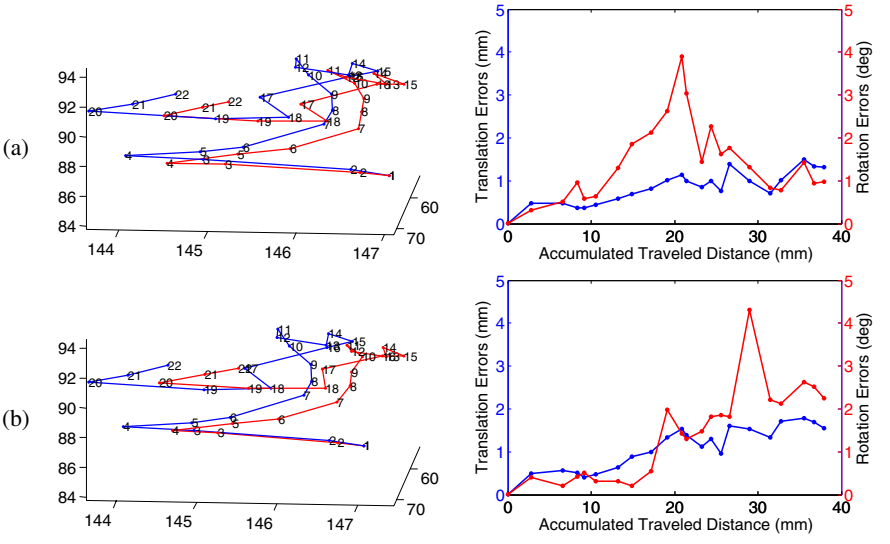


Fig. 5. Results from the video segment 1 with Method I (a) and Method II (b). Left:The trajectories of our method (red) and Optotrak (blue), Right:frame to frame relative error along the trajectories.

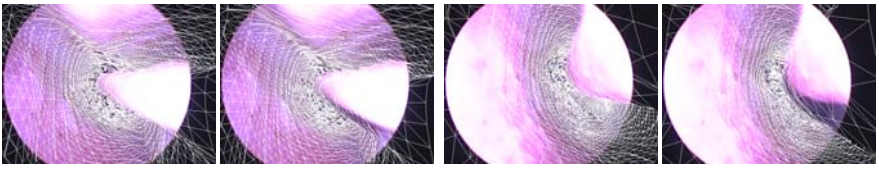


Fig. 6. Comparison of our method (Left) and Optotrak (Right). Left Pair, beginning of tracking. Right Pair, end of tracking.

4 Discussion and Conclusion

Our results indicate that video-CT registration can be computed to within $1mm$ of our ground truth validation system. Recall $1mm$ is the expected error of the Optotrak measurements alone. We hypothesize that the video-CT solution is well below $1mm$ error since we are within the error envelope of the Optotrak, and because visual inspection of the registrations often shows a more consistent result from the video-CT algorithm. We present relative error as we continue investigate a method that is more accurate than the Optotrak tracking a rigid endoscope.

In future work, we will validate that we are indeed below $1mm$ error by using CT-compatible screws that would be imaged in both the endoscope and CT. Using the segmented location from both the endoscope and CT, a 2D-3D pose estimator can be used to validate the registration error to within the resolution of the CT image itself.

We will also investigate robust features to more accurately track over large translations. While SIFT features do offer image features to track, they do not track well over large translation. SIFT features fail to track over large translation in endoscopy because SIFT is not particularly robust to illumination changes. In endoscopy the light source is collocated with the camera and thus the illumination is always changing as the camera moves. For a more accurate reconstruction a large translation is preferred. Beyond image features, the selection of image pairs to use for reconstruction can be automated. It is important to select image pairs with motion greater than the noise of the image features tracked to ensure accurate reconstruction.

The current system is implemented in Matlab and processes the data offline taking about two minutes per frame pair. We focused on accuracy and robustness instead of speed. This methodology is common to vision literature. With further engineering the algorithm could be turned into an online system suitable for clinical use.

We also acknowledge that our algorithm does require that there be enough 3D structure and 2D texture in the scene to register and to track. If the endoscope is pointed at a flat surface or in a textureless region, our algorithm would not perform well. However, our algorithm could be used for local registration enhancement and would therefore add higher accuracy capability to existing tracking systems. We envision a visual re-registration feature that would offer surgeons the option to have higher accuracy for the current scene.

Acknowledgments. This work was supported principally by the National Institutes of Health under grant number R21 EB005201. Further support was provided by the NSF ERC Grant EEC9731748 and by Johns Hopkins University. We thank Dr. Darius Burschka for providing the porcine video data.

References

1. Nasseri, S.S., Kasperbauer, J.L., Strome, S.E., McCaffrey, T.V., Atkinson, J.L., Meyer, F.B.: Endoscopic transnasal pituitary surgery: Report on 180 cases. American Journal of Rhinology 15(4), 281–287 (2001)
2. Chassat, F., Lavallée, S.: Experimental protocol of accuracy evaluation of 6-D localizers for computer-integrated surgery: Application to four optical localizers. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 277–284. Springer, Heidelberg (1998)
3. Shahidi, R., Bax, M., Maurer, C.R., Johnson, J., Wilkinson, E., Wang, B., West, J., Citardi, M., Manwaring, K., Khadem, R.: Implementation, calibration and accuracy testing of an image-enhanced endoscopy system. Med. Imag., IEEE Trans. 21(12), 1524–1535 (2002)
4. Lapeer, R., Chen, M.S., Gonzalez, G., Linney, A., Alusi, G.: Image-enhanced surgical navigation for endoscopic sinus surgery: evaluating calibration, registration and tracking. Intl. J. of Med. Rob. and Comp. Assisted Surg. 4(1), 32–45 (2008)
5. Helferty, J.P., Hoffman, E.A., McLennan, G., Higgins, W.E.: CT-video registration accuracy for virtual guidance of bronchoscopy. In: SPIE Med. Imaging, vol. 5369, pp. 150–164 (2004)
6. Burschka, D., Li, M., Ishii, M., Taylor, R.H., Hager, G.D.: Scale-invariant registration of monocular endoscopic images to CT-scans for sinus surgery. MIA 9, 413–426 (2005)
7. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 20, 91–110 (2003)

8. Vedaldi, A.: SIFT for matlab, http://www.vlfeat.org/~vedaldi/code/sift.html (last accessed May 29, 2009)
9. Delponte, E., Isgr, F., Odone, F., Verri, A.: SVD-matching using SIFT features. In: Proc. of the Intl. Conf. on Vision, Video and Graphics, July 2005, pp. 125–132 (2005)
10. Wang, H., Mirota, D., Ishii, M., Hager, G.: Robust motion estimation and structure recovery from endoscopic image sequences with an Adaptive Scale Kernel Consensus estimator. In: CVPR. IEEE Conf. on, June 2008, pp. 1–7 (2008)
11. Nister, D.: An efficient solution to the five-point relative pose problem. IEEE Trans. PAMI 26(6), 756–770 (2004)
12. Mirota, D., Taylor, R.H., Ishii, M., Hager, G.D.: Direct endoscopic video registration for sinus surgery. In: Medical Imaging 2009: Visualization, Image-guided Procedures and Modeling, Proc. of the SPIE, February 2009, vol. 7261, pp. 72612K–1,72612K–8 (2009)
13. Chetverikov, D., Svirko, D., Stepanov, D., Krsek, P.: The trimmed iterative closest point algorithm. ICPR 3, 545–548 (2002)
14. Du, S., Zheng, N., Ying, S., You, Q., Wu, Y.: An extension of the ICP algorithm considering scale factor. ICIP 5, 193–196 (2007)
15. Lu, C.P., Hager, G., Mjolsness, E.: Fast and globally convergent pose estimation from video images. PAMI, IEEE Trans. 22(6), 610–622 (2000)
16. Bouget, J.Y.: The matlab camera calibration toolkit, http://www.vision.caltech.edu/bouguetj/calib_doc/ (last accessed, March 3 2008)
17. Park, F., Martin, B.: Robot sensor calibration: solving AX=XB on the Euclidean group. IEEE Transactions on Robotics and Automation 10(5), 717–721 (1994)

Correcting Motion Artifacts in Retinal Spectral Domain Optical Coherence Tomography via Image Registration

Susanna Ricco[1], Mei Chen[2], Hiroshi Ishikawa[3,4], Gadi Wollstein[3], and Joel Schuman[3,4]

[1] Department of Computer Science, Duke University, Durham, NC
sricco@cs.duke.edu
[2] Intel Research Pittsburgh, Pittsburgh, PA
[3] UPMC Eye Center, University of Pittsburgh Medical Center, Pittsburgh, PA
[4] Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA

Abstract. Spectral domain optical coherence tomography (SD-OCT) is an important tool for the diagnosis of various retinal diseases. The measurements available from SD-OCT volumes can be used to detect structural changes in glaucoma patients before the resulting vision loss becomes noticeable. Eye movement during the imaging process corrupts the data, making measurements unreliable. We propose a method to correct for transverse motion artifacts in SD-OCT volumes after scan acquisition by registering the volume to an instantaneous, and therefore artifact-free, reference image. Our procedure corrects for smooth deformations resulting from ocular tremor and drift as well as the abrupt discontinuities in vessels resulting from microsaccades. We test our performance on 48 scans of healthy eyes and 116 scans of glaucomatous eyes, improving scan quality in 96% of healthy and 73% of glaucomatous eyes.

1 Introduction

Many ophthalmologists use optical coherence tomography (OCT) to improve diagnosis of retinal diseases. OCT generates cross-sectional and 3D images of retinal structure by detecting the magnitude and echo time delay of light. The newest version of OCT, spectral domain OCT (SD-OCT) [1], detects all the light echoes in a single axial scan (A-scan) in parallel. Cross-sectional images (B-scans) are created by acquiring a series of A-scans as the OCT beam is scanned in the transverse direction. SD-OCT can image the three-dimensional retinal structure with an axial resolution of up to $\approx 2\mu$m and an imaging speed of up to 50,000 A-scans per second.

An important application of retinal SD-OCT is the early diagnosis of glaucoma, a progressive blinding disease. Early intervention has been shown to slow progression [2], but early detection of the disease is challenging. The precise structural measurements available from SD-OCT volumes (such as retinal nerve fiber layer thickness maps [3]) can be used to detect structural changes before they result in noticeable vision loss.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 100–107, 2009.

A typical SD-OCT scan pattern contains 40,000 A-scans and takes between 1 and 1.5 seconds to complete. Any eye movement during imaging will corrupt the data, making structural measurements and glaucoma diagnosis unreliable. Eye movement during imaging is reduced by asking the subject to fixate on a target; however, our eyes remain in motion even during conscious fixation, undergoing *fixational eye movements* [4]. Fixational eye movements encompass three different types of motion: ocular tremor, ocular drift, and microsaccades. Microsaccades, which introduce the most severe motion artifacts, move the fixated point around 30 arcminutes almost instantaneously and occur roughly once per second in normal individuals.

Figure 1(a) shows a two-dimensional projection of an SD-OCT volume containing motion artifacts. Here, the axial dimension has been collapsed by summing voxel intensities, resulting in a view of the eye (the *en face* image) analogous to a traditional fundus image or other 2D imaging techniques. Note the discontinuity in the vessels pointed out in the bottom half of the image. These discontinuities indicate that microsaccades occurred during scan acquisition.

Coupling the imager to a tracker [5] results in an improvement in the OCT data, but with high production costs and increased scanning time.

We take a complementary approach to eliminating transverse motion artifacts: we correct motion artifacts after imaging is complete via image registration. Immediately after capturing the SD-OCT volume, the OCT system captures an scanning laser ophthalmoscopy (SLO) image of the retina (Fig. 1(d)). Because capture of the SLO image is virtually instantaneous relative to time required for OCT capture, we can treat the SLO image as an artifact-free reference image. The OCT *en face* image should match the SLO image if the scan is also artifact-free. Any structural misalignment indicates that the OCT volume contains motion artifacts and does not accurately reflect the subject's anatomy. We remove these artifacts by warping the volume to make the corresponding *en face* image match the SLO image as closely as possible.

Prior work on correcting motion artifacts in OCT volumes without requiring active tracking registers consecutive B-scans by finding the shifts in the $Z-$ and $X-$dimensions that maximize cross-correlation between adjacent B-scans [6], correcting transverse and axial artifacts resulting from ocular drift. While our method does not correct axial artifacts, we surpass this method in correcting transverse artifacts by allowing for more complex motions, including fast motion during a single B-scan and motion in the $Y-$dimension.

2 Methods

We remove motion artifacts by registering the *en face* image constructed from an SD-OCT volume to the SLO reference image and applying the resulting warp to the volume at each depth. Registration proceeds in two steps. The first step corrects for artifacts from tremor and drift; the last step, the main contribution of our technique, corrects vessel discontinuities caused by microsaccades. Results from each correction step on an example scan are shown in Fig. 1.

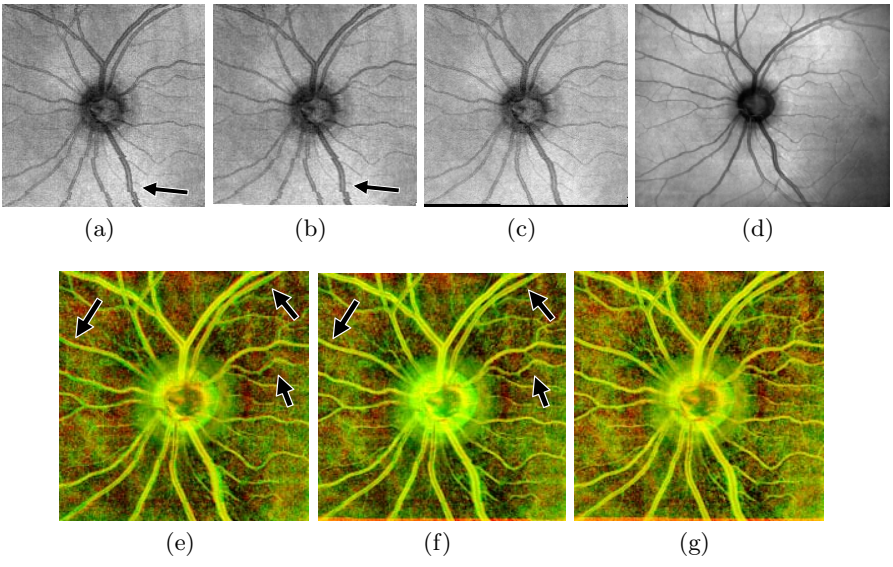


Fig. 1. Results of motion artifact correction on an example scan. The *en face* image from the uncorrected scan (a) shows vessel discontinuities (arrow) resulting from microsaccades. Correcting for tremor and drift (b) fails to correct these discontinuities, but microsaccade correction does (c), matching the SLO reference image (d). Images (e)–(g) show the *en face* image in the red channel and the SLO image in the green channel before correction, after correcting tremor and drift, and after full correction, respectively. Correctly aligned regions appear yellow; arrows point to improved alignment after drift correction.

2.1 Detecting Vessels

At each registration step, we focus on matching the location of vessels, the main features in retinal images. Our vessel detection algorithm is a simplified implementation of the technique proposed by Lam and Yan [7], which uses hysteresis thresholding to find ridges in the divergence of the image gradient. After scaling the computed divergence to lie between 0 and 1, we use upper and lower thresholds of 0.2 and 0.075, respectively, and discard regions containing fewer than five pixels above the upper threshold.

2.2 Correcting Tremor and Drift

We correct for tremor and drift using the elastic registration technique of Periaswamy and Farid [8]. Motion between the two images is described by an affine transformation that may vary across the image but is assumed to be constant within a local window, Ω. Specifically, we solve for the parameter vector $\boldsymbol{m}$ that solves

$$m_7 I^{\mathrm{SLO}}(x, y) + m_8 = I^{\mathrm{OCT}}(m_1 x + m_2 y + m_5, m_3 x + m_4 y + m_6) \qquad (1)$$

for all locations $(x, y) \in \Omega$ in the least squares sense. We define our window to capture temporally distinct phenomena, including pixels corresponding to A-scans taken within a time interval τ. We use the same window for every A-scan in a given B-scan, drastically reducing the amount of computation required.

It is customary in motion estimation applications to estimate motion on a scale pyramid. Because tremor and drift are both small amplitude motions resulting in relatively small errors, we omit this coarse-to-fine scheme, eliminating the need to smooth the images and recompute gradients at each level of the pyramid. Instead, we vary the size of our window. This allows us to first compute a global deformation that brings the two images mostly into alignment before focusing on more localized motion artifacts.

After correcting for drift and tremor, we are left with *en face* images that are well aligned with the SLO reference image except in places where a microsaccade occurred. The result of tremor and drift correction is shown in Fig. 1(b). Note that the drift correction has improved the alignment of most vessels (correctly overlapping areas appear yellow in Fig. 1(f)), particularly in the top half of the image, but the vessel discontinuities remain.

2.3 Correcting Microsaccades

The main contribution of our algorithm is the ability to correct vessel discontinuities that result from primarily horizontal microsaccades. At a microsaccade, the location of the eye between subsequent B-scans (and between different A-scans within a single B-scan) changes significantly, causing the location of vessels in the image to shift abruptly between two adjacent rows. The necessary correction before and after the time of the microsaccade is drastically different, violating the local smoothness assumption made by many motion estimation techniques.

We correct microsaccades by finding the horizontal shift at each pixel in the scan that best aligns the result of tremor and drift correction with the SLO image. We consider the results of the earlier correction and the SLO image as signals in the time domain. We then find the least cost set of feasible correspondences between the two signals. This procedure is commonly referred to as *dynamic time warping* [9] in the signal processing community, and is also used on corresponding scanlines to find dense disparity maps in stereo vision [10]. This procedure corrects horizontal microsaccades but ignores vertical motion.

Our cost function penalizes matches based on the squared difference in pixel intensities. To force vessels to align properly, the intensity of pixels inside vessels should be as different as possible from pixels outside the vessels. We also want the non-vessel tissue to be as uniform as possible to avoid detecting spurious saccadic events. As vessel detection is rarely perfect, we do not rely solely on a binary mask. Instead, we apply a scheme that darkens pixels within vessels and normalizes the intensity of non-vessel tissue while tolerating errors in vessel detection. We replace the intensity at pixels detected as vessels with the result of adaptive histogram equalization. This darkens vessels but does not drastically alter the intensity of pixels incorrectly detected as vessels. We also apply a linear transformation to the intensities of the remaining pixels so that

a line fit to the intensities on each row is horizontal, with corresponding rows in the SLO and *en face* images sharing the same mean and standard deviation in intensities.

After adjusting pixel intensities, we convert the images to one-dimensional signals and find the lowest cost set of correspondences using dynamic programming. The total cost is equal to the sum of the squared differences in intensity for each pair of matched pixels plus a constant P (we use twice the standard deviation in pixel intensity) for each unmatched pixel. The returned correspondences are converted to horizontal shifts and used to warp the partially corrected *en face* image. Finally, an additional iteration of tremor and drift correction removes residual vertical motion artifacts, which could not be corrected on the first pass due to the presence of significant saccades. The fully corrected *en face* image for the example from Fig. 1(a) is shown in Fig. 1(c).

2.4 Preserving Optic Disc Contour

Microsaccade correction is sensitive to vessels that appear darker than the surrounding tissue in the *en face* image but brighter than the surrounding tissue in the SLO image. This is particularly common inside the optic disc and can cause the optic disc contour to become distorted after motion correction. Because the optic disc contour is an important feature in glaucoma diagnosis, our correction must never introduce errors in this region. We therefore disallow horizontal shifts in the region surrounding the optic disc, which is automatically detected by thresholding the distance to the top surface of the retina in the volume.

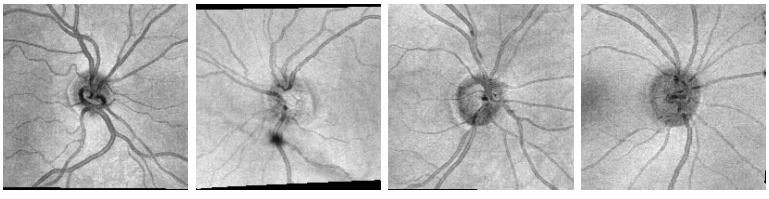
3 Experimental Results

We implemented our algorithm in Matlab and tested it on 164 SD-OCT scans: 48 from healthy eyes and 116 from glaucomatous eyes. There are 24 healthy (the H-D set) and 58 glaucomatous examples (the G-D set) containing noticeable vessel discontinuities from microsaccades. The remaining examples (abbreviated as the H-ND and G-ND sets) contain drift and tremor artifacts but no noticeable microsaccades. Figure 2 shows *en face* images from each of the four sets before and after correction. Each volume is $200 \times 200 \times 1024$ voxels and takes at most 50 seconds to correct on a desktop PC (Intel Pentium 3.2GHz, 2GB RAM).

3.1 Qualitative Assessment

Two OCT experts independently compared the quality of each scan before and after correction. They evaluated both the overall fit to the SLO reference image (reflecting correction of tremor and drift artifacts) and the number of noticeable vessel discontinuities (reflecting correction of microsaccades), classifying the correction as either improved, unchanged, or degraded in both areas. The two experts' evaluations were combined into a consensus evaluation for each scan. For

(a) H-ND before (b) H-D before (c) G-ND before (d) G-D before



(e) H-ND after (f) H-D after (g) G-ND after (h) G-D after

Fig. 2. Results of motion artifact correction on example scans from our four datasets (abbreviations from Sect. 3). *En face* images for each example are shown both before and after correction. Arrows point to vessel discontinuities before correction.

the corrected scan to be considered an improvement, at least one expert must have rated the output as improved and neither can have rated it as degraded.

The results of the experts' assessment are summarized in Table 1. Overall fit improved in 45 of the 48 of the healthy examples, remaining unchanged in the other three. In the glaucomatous examples, the overall fit improved in 96 of the 116 examples, remaining unchanged in 17 and deteriorating in 3. Of the 24 H-D examples, we improve vessel continuity in 19, leaving the remaining five unchanged. We also improve vessel continuity in 45 of the 58 G-D examples, leaving 3 unchanged and causing deterioration in the remaining 10. Ultimately, we consider the total quality of a scan to be improved if either overall fit or vessel continuity improved without either deteriorating. Our correction improved the total quality of 46 of 48 healthy (96%) and 85 of 116 glaucomatous (73%) scans.

Table 1. A summary of the qualitative evaluation of our performance. Experts assessed improvement after correction in overall alignment (O) and vessel continuity (V). Below, + indicates improvement, − indicates degradation, and = indicates no change detected. The three bold columns are the evaluations considered to be improved total scan quality.

Dataset	**O+ V+**	**O+ V=**	**O= V+**	O= V=	O+ V−	O− V+	O= V−	O− V=	O− V−
H-ND	**0**	**22**	**0**	1	1	0	0	0	0
H-D	**17**	**5**	**2**	0	0	0	0	0	0
G-ND	**11**	**27**	**0**	6	7	0	5	1	1
G-D	**43**	**2**	**2**	1	6	0	3	0	1

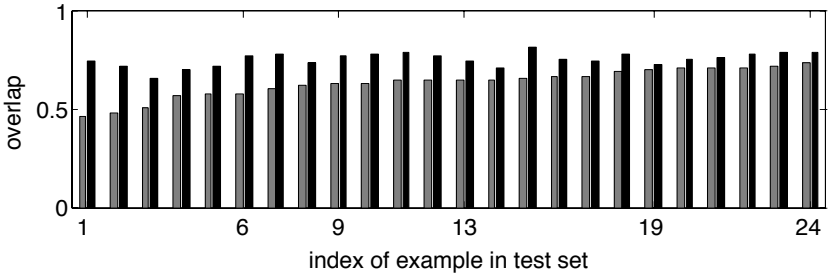


Fig. 3. Percent overlap between corresponding vessels in *en face* image and SLO reference image before and after correction for the H-D set. Gray bars show overlap before correction; black bars show overlap after motion artifact correction. Examples are sorted in order of increasing overlap before correction. Examples 9 and 19 are those judged by the experts to have no improvement in overall alignment after correction. Example 6 is shown in Fig. 2(b) and (f). Example 13 is shown in Fig. 1.

3.2 Quantitative Assessment

We evaluate the performance quantitatively by computing the improvement in overlap between vessels in the SLO image and vessels in the motion corrected *en face* image. This single measurement combines the contributions of both tremor and drift correction and the correction of any microsaccades. We define overlap as the fraction of pixels belonging to each major vessel in the *en face* image that also belong to the corresponding vessel in the SLO image (based on manually labeled vessels). We use this variant of the Jaccard index to give a value of 1 for vessels which are correctly aligned but appear thinner in the *en face* image than in the SLO image due to the different imaging protocols.

Figure 3 shows the overlap improvement on the scans from the H-D set. In every case, the overlap is greater after correction. The mean overlap before correction over these 24 examples is 0.6348, which increases to a mean of 0.7522 after artifact removal. For the other three sets of scans the mean overlap increases from 0.6083 to 0.7403 (G-D), from 0.6396 to 0.7168 (G-ND), and from 0.6486 to 0.7469 (H-ND). Overlap decreases in only five examples, all in the G-ND set.

4 Discussion

In this paper, we described a system that automatically corrects transverse motion artifacts in retinal SD-OCT after scan acquisition. Both qualitative and quantitative evaluations show that our system improves the quality of the majority of scans, even in subjects with retinal diseases.

The main contribution of our technique is the ability to correct for artifacts resulting from microsaccades. We use a more flexible motion model than previous methods, allowing a different correction for individual A-scans within a single B-scan. Ophthalmologists use vessel continuity as a key measure of scan quality

and often discard scans containing discontinuities larger than the vessel diameter. With our system, the majority of these previously discarded scans are now useful.

Limitations that remain to be addressed include our handling of the optic disc and artifacts from vertical saccades. Because we disallow horizontal shifts within the optic disc, we cannot correct saccades occurring at that point in the scan. Applying more sophisticated shape analysis to the optic disc may allow for saccade correction without distorting the optic disc contour. Although we do correct smooth vertical motions, we do not specifically address vertical saccades. It may be possible to consider the horizontal and vertical components of saccades independently and correct each with our current saccade correction technique.

Acknowledgments. The first author performed this work while at Intel Research Pittsburgh; the research was funded by Intel Research Pittsburgh and the University of Pittsburgh Medical Center.

References

1. Wojtkowski, M., Srinivasan, V., Fujimoto, J., Ko, T., Schuman, J., Kowalczyk, A., Duker, J.: Three-dimensional retinal imaging with high-speed ultrahigh-resolution optical coherence tomography. Ophthalmology 112(10) (2005)
2. Heijl, A., Leske, M., Bengtsson, B., Hyman, L., Bengtsson, B., Hussein, M.: Reduction of intraocular pressure and glaucoma progression results from the Early Manifest Glaucoma Trial. Archives of Ophthalmology 120(10) (2002)
3. Townsend, K., Wollstein, G., Schuman, J.: Imaging of the retinal nerve fibre layer for glaucoma. British Journal of Ophthalmology 93(2) (2009)
4. Martinez-Conde, S., Macknik, S., Hubel, D.: The role of fixational eye movements in visual perception. Nature Reviews Neuroscience 5(3) (2004)
5. Hammer, D., Ferguson, R., Magill, J., Paunescu, L., Beaton, S., Ishikawa, H., Wollstein, G., Schuman, J.: Active retinal tracker for clinical optical coherence tomography systems. Journal of Biomedical Optics 10 (2005)
6. Zawadzki, R., Fuller, A., Choi, S., Wiley, D., Hamann, B., Werner, J.: Correction of motion artifacts and scanning beam distortions in 3D ophthalmic optical coherence tomography imaging. Ophthalmic Technologies XVII 6426(1) (2007)
7. Lam, B., Yan, H.: A Novel Vessel Segmentation Algorithm for Pathological Retina Images Based on the Divergence of Vector Fields. IEEE Transactions on Medical Imaging 27(2) (2008)
8. Periaswamy, S., Farid, H.: Elastic registration in the presence of intensity variations. IEEE Transactions on Medical Imaging 22(7) (2003)
9. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Transactions on Acoustics, Speech and Signal Processing 26(1) (1978)
10. Birchfield, S., Tomasi, C.: Depth Discontinuities by Pixel-to-Pixel Stereo. International Journal of Computer Vision 35(3) (1999)

Single Fiber Optical Coherence Tomography Microsurgical Instruments for Computer and Robot-Assisted Retinal Surgery

Marcin Balicki[1,*], Jae-Ho Han[1], Iulian Iordachita[1],
Peter Gehlbach[2], James Handa[2], Russell Taylor[1], and Jin Kang[1]

[1] ERC for Computer Integrated Surgery, Johns Hopkins University
[2] Wilmer Eye Institute, Johns Hopkins Medical, Baltimore, MD, USA
{marcin,iordachita,rht}@jhu.edu

Abstract. We present initial prototype and preliminary experimental demonstration of a new class of microsurgical instruments that incorporate common path optical coherence tomography (CP-OCT) capabilities. These instruments may be used freehand or with robotic assistance. We describe a prototype 25 gauge microsurgical pick incorporating a single 125 μm diameter optical fiber interfaced to a Fourier Domain CP-OCT system developed in our laboratory. For initial experimentation, we have interfaced this instrument with an extremely precise, cooperatively controlled robot. We describe the tool, system design, and demonstration of three control methods on simple phantom models: 1) enforcement of safety constraints preventing unintentional collisions of the instrument with the retinal surface; 2) the ability to scan the probe across a surface while maintaining a constant distance offset; and 3) the ability to place the pick over a subsurface target identified in a scan and then penetrate the surface to hit the target.

1 Introduction

Vitreoretinal surgery addresses prevalent sight-threatening conditions such as retinal detachment, macular pucker, macular holes, and conditions in which epiretinal scar tissue is removed. The technical demands placed on the surgeon by these procedures are extreme. In current practice, retinal surgery is performed under an operating microscope with free-hand instrumentation. Human limitations include an inability to clearly visualize surgical targets, physiological hand tremor, and lack of tactile feedback in tool-to-tissue interactions. In addition, tool limitations, such as lack of proximity sensing or smart functions, are important factors that contribute to surgical risk and reduce the likelihood of achieving surgical goals. Current instruments do not provide physiological or even basic interpretive information, e.g. the distance of the instrument from the retinal surface, the depth of instrument penetration into the retina or an indication of the force exerted by the instrument on the retinal tissues. Surgical outcomes (both success and failure) are limited, in part, by technical hurdles that cannot be overcome by current instrumentation. For example, in the most technically

[*] Supported in part by the NSF EEC9731748, NIH 1R01 EB007969-01 and ARCS Foundation.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 108–115, 2009.

demanding cases, there may not be a set of tools that allows the "typical" retina surgeon to remove sufficient epiretinal scar tissue to ensure surgical success.

In this paper, we will use the peeling of epiretinal membranes (ERMs) from the surface of the retina as a motivating example. ERM peeling is a common and extremely demanding surgical procedure. ERMs are composed of scar tissue that forms on the surface of the retina, contracts and compromises retinal function. ERMs are present in 2-6.4% of people [1]. Visual dysfunction resulting from ERM includes: blurred vision, image distortion, and altered image size. Surgical removal of an ERM involves identifying or creating an "edge" that is then grasped and peeled. Some ERMs provide clear visual evidence of edges that may be grasped. Others require creation of an edge by the surgeon. This may be performed by incising the membrane surface, by bluntly creating an edge, or by gently grasping the membrane with a forceps and creating a tear in the ERM. Each of these maneuvers requires excellent visualization, high levels of manual dexterity and micro-instrumentation. Furthermore, this procedure is performed with a comparatively large metal instrument without tactile sensation. During this time, a patient's involuntary and voluntary movement must be manually compensated for by the surgeon while the instrument is in direct contact with fragile intraocular tissue. Incorrect micron-magnitude movements can cause retinal tears, retinal detachment, visual field defects, retinal hemorrhage, local retinal edema, nerve fiber layer injury, and macular holes, all of which can contribute to blindness.

Optical Coherence Tomography (OCT) - provides very high resolution (micron scale) images of anatomical structures within the tissue. Within Ophthalmology, OCT systems typically perform imaging through microscope optics to provide 2D cross-sectional images ("B-mode") of the retina. These systems are predominantly used for diagnosis, treatment planning and in a few cases, for optical biopsy and image guided laser surgery [3-6].

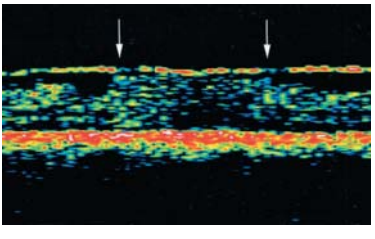


Fig. 1. Preoperative OCT scan of multiple points of retinal attachment (arrows) by epiretinal membrane [2]

Fig. 1 shows a typical OCT scan of an epiretinal membrane demonstrating multiple points of retinal attachment (arrows). ERMs are recognized by OCT as thin, highly reflective bands anterior to the retina. A potential dissection plane between the ERM and the retina is clearly visible in the scan, but is invisible to the surgeon through an operating microscope, even with very high magnification. In other work [7] our laboratory has explored registration of preoperative OCT images to intraoperative microscope images to aid in identifying ERM edges for initiating ERM removal. However, ERMs can grow and further distort retinal architecture. It is therefore, unclear whether preoperative images would provide a useful guide if the interval between the preoperative image acquisition and surgery allows for advancement of the ERM. Direct imaging of the ERM relative to the surgeon's instruments would be very useful either as a replacement for or supplement to the preoperative OCT images. In addition, direct imaging of the local anatomy relative to the instruments can be used to provide feedback on tool-to-tissue distances and real-time updates during tissue dissection.

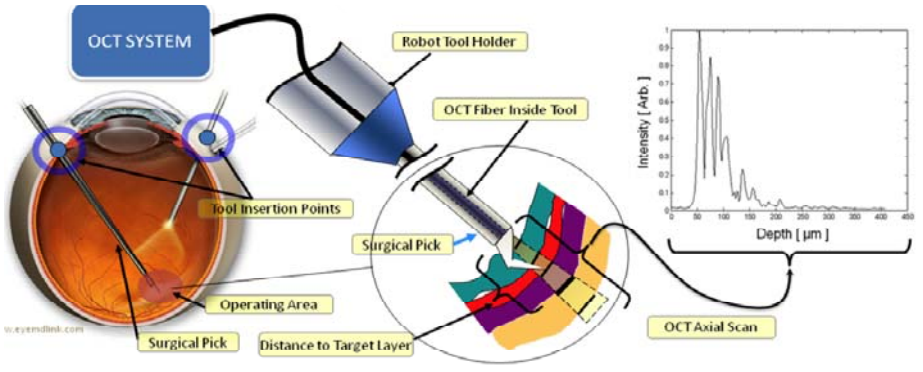


Fig. 2. Left: Proposed use of the instrument inside the human eye. Right: Axial OCT scan.

These considerations have led us to explore development of sub-millimeter diameter microsurgical instruments incorporating OCT sensing capability for directly sensing tissue planes beyond the instrument tip. Our approach (illustrated in Fig. 2) relies on the use of common path OCT (CP-OCT) sensing through an optical fiber built into the instrument shaft. Compared to alternative approaches, CP-OCT is simple, robust, and affordable [8], and also permits an extremely compact tool design. This "A-mode" sensing capability can either be used with conventional free-hand instruments or (as here) can be combined with a robotic microsurgery platform such as the JHU EyeRobot [9] or the CMU Micron [10]. When integrated with a robot, this instrument may be scanned to produce images of the local anatomy or integrated in various ways to the control the robot. This integration will allow design of a system that will have safety barriers that prevent collision with the retina, be able to maintain a constant distance from the retina during scanning, and be able to promote accurate targeting of anatomic features within the retina.

In this paper, we demonstrate integration of a single fiber OCT sensor into a 0.5 mm diameter microsurgical instrument designed for vitreoretinal surgical interventions. The instrument position is inherently registered to OCT by the fact that the surgical tip is visible in the OCT image. We have integrated this system into a robotic platform in order to investigate various control methods for its use in robotically assisted microsurgery.

2 Materials

OCT Integrated Pick Instrument. Vitreoretinal picks are commercially available in a variety of sizes with the most commonly used ranging from 25-20 gauge (0.5 mm to 0.9 mm diameter). Sharp picks are used to incise taut surface membranes and engage and elevate membrane edges for ERM peeling. The simplicity of this basic tool permits the surgeon to quickly craft a sharp edged pick from an appropriately sized surgical needle by bending the beveled tip with a needle holder. The surgeon co-authors routinely use this method to create such picks.

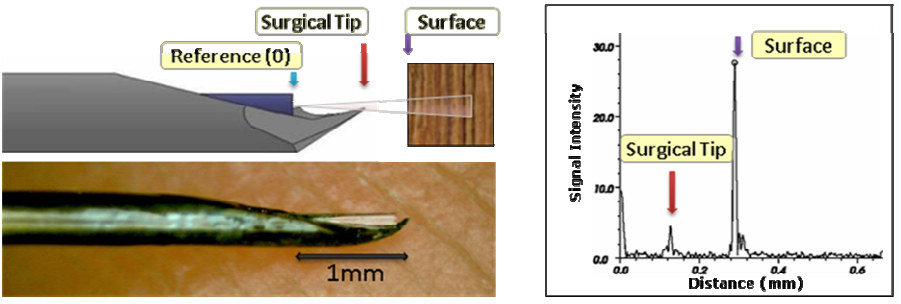


Fig. 3. Top Left: CAD side view of microsurgical pick with integrated fiber optic OCT probe. Bottom Left: Actual prototype. Right: A-Scan image of a sample.

For our experiments, we have adapted this method to permit incorporation of a single fiber OCT probe to enable simultaneous A-mode imaging and tissue manipulation. Our prototype picks were constructed from 25 gauge, 38mm surgical needles with 0.24mm ID and 0.5mm OD. We bent the beveled point approximately 200-300µm from the tip so that the tip intersected with the central axis of the needle lumen. A cleaved optical fiber stripped to glass cladding was inserted through the lumen, bonded in place approximately 135µm from the tip, and interfaced to the CP-OCT system. The tip is visible in the OCT A-mode image, thus providing a reference point for the relative distance of the tool to tissue being scanned (Fig. 3).

Optical Coherence Tomography System. The optical fiber from the tool is integrated with a common path Fourier domain OCT system developed in our laboratory and is fully described in [8]. It is the simplest, most compact imaging technique of its kind. This system is robust, inexpensive, and can utilize simple interchangeable probes. Our system uses an SLED light source centered near 0.8-µm wavelength with a fiber optic directional 2x2 coupler, and a USB spectrometer (HR4000, Ocean Optics Inc.). The optical fiber probes presented here are a single mode fiber (NA 0.14) with ~9µm core size, 125µm cladding diameter and ~245µm diameter outer polymer buffer. The imaging depth is 1-1.5 mm in tissue and somewhat less in our wax sample materials. In air, the OCT beam diverges at an angle of 16 degrees and effective axial resolution is ~8.8µm. In water or tissue the divergence is ~12 degrees with ~6.6 µm resolution. A sample axial image (A-Mode) with zero reference point located at the tip of the fiber is shown in Fig. 2. By moving the probe laterally, a sample cross section image is generated (B-Mode).

The CP-OCT system provided the robot with the distance from the probe to the surface of the sample. Each axial scan was processed by first thresholding to remove baseline noise, and then smoothing with a Savitzky-Golay filter to preserve peak locations, amplitudes and widths. To locate the peaks, a simple zero-crossing of signal derivative was used. The acquisition time and signal processing yields a sampling rate of ~200Hz with approximate latency of one frame.

Sample materials. We have developed two artificial phantoms for consistent evaluation in these preliminary experiments. For the safety barrier and surface tracking tasks, we use a composite of three 60µm thick layers of scotch tape on a dense

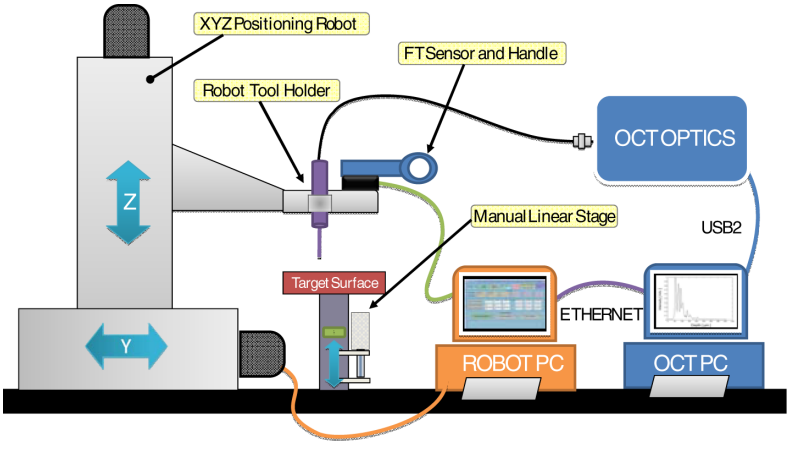


Fig. 4. Experimental system overview

wooden plank. This provides a strong multi-peak axial scan signal that is analogous to that generated by the multilayered structures of the retina. For the targeting tasks, we needed a phantom with 100-300µm diameter cavities located near the surface and the ability to display any physical interaction with the sample. Sheet lining wax (McMaster 8691K31) with a low softening point (135°C) was placed on an aluminum sheet and heated to boiling. Rapid cooling produces many thin-walled bubbles and any physical contact between the pick and sample surface leaves permanent marks visible in OCT.

Experimental Setup. Our experimental setup is illustrated in Fig. 4. The microsurgical instrument is mounted on an extremely precise Cartesian positioning robot consisting of three orthogonally mounted linear stages. The position resolution is 1µm and measured repeatability is about ± 1 µm for range of motion required for this experiment (<2mm). This robot was chosen for experimental convenience, and would be replaced in actual practice by a system similar to [9-10]. The robot is interfaced to a PC workstation through a commercial motion controller (Galil DMC 1886). A commercial 6 DOF force-torque sensor (ATI Nano43) is mounted on the robot and also interfaced to the PC as a user interface. Open-source robot control software developed in our laboratory [11] is used to implement cooperative "hands on" control for the robot's motion. The OCT system is implemented on a separate PC and communicates to the robot PC via Ethernet. The test samples are mounted on a separate manually actuated micrometer stage placed beneath the probe. An inspection video microscope (Omano E-ZVU/V15) is positioned to simultaneously image the side view of the sample and probe at 90X magnification (not shown).

3 Methods

We have demonstrated our system in the context of three sample tasks: enforcement of safety barriers, "tracking" to maintain a constant distance from a surface, and accurate placement of the probe on targets identified in a scanned OCT image. Relatively

low velocities were chosen based on surgical preference when operating close to the retina and the limited movement in the eye during surgery due to constraints by insertion of tools through the sclera and effects of anesthetics.

In the **safety barrier** task, the system enforced a safety constraint to prevent the probe from approaching the target surface closer than a specified threshold distance. The robot moved freely within the 1D workspace to comply with forces exerted by the user on the control handle, with the exception of the forbidden boundary sensed via the OCT. This "virtual wall" was reached when the tip of the probe was located ~150µm from the sample surface. A bare optical fiber was used as a probe. Five trials were performed with different robot velocity limits: 100, 200, 300, 400, 500 µm/sec.

In the **surface tracking** task, the sample surface was moved up and down with the manual micrometer stage while the robot was controlled to maintain a constant OCT-reported distance of 150 µm from the sample surface. One intention for the surface tracking was to correct for retinal motion due to respiratory function, hence we chose the sinusoidal frequency to be around 0.2 Hz and magnitude that encompasses expected ranges of retinal motion.

In the **targeting** task, the robot was instructed to scan the pick in 2µm increments laterally across the sample surface. The A-mode images from the OCT system were then combined to make a B-mode image. The evolving B-mode image was displayed continuously to the user, who could use a mouse at any time to click on a target within the image. The robot would then interrupt its scan, move back to a position over the identified target, and then slowly insert the pick tip to the depth of the designated target, based on depth feedback provided by the OCT system. The probe was then withdrawn and a second B-mode scan was taken to observe the created defect in the sample surface.

4 Results

The results for the safety barrier task are shown in Fig. 5(left). The observed overshoot into the "unsafe" zone was negligible for robot speeds up to ~300 µm/sec and still quite modest up to ~500 µm/sec, using only a very naïve control method. Further improvements are expected with our next generation OCT system (which has higher sample rate) and with refinements in the robot design and control.

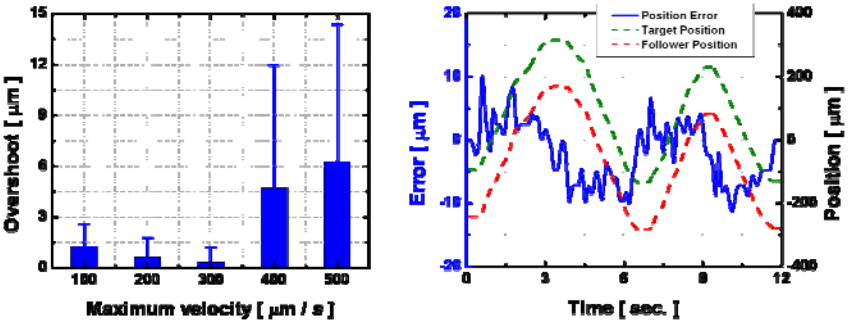


Fig. 5. Results: Left: Safety Barrier - Overshoot error vs. maximum allowable velocity. Right: Surface tracking with 150µm standoff.

The results of dynamic response when tracking sinusoidal surface motion are shown in Fig. 5(right). We were able to keep the tool tip within about 10 µm of the desired 150 µm standoff from the target surface while the surface moved at about 200 µm/sec.

The results of the targeting task are illustrated in Fig. 6. The top B-Scan shows a subsurface bubble with 30-50µm membrane and the user specified incision location. The bottom shows the post-intervention B-Scan with an overlay depicting approximate orientation and location of the instrument tip at the target. The defect clearly resembles the geometry of the tip, as well as good alignment with planned incision position.

5 Discussion and Future Work

"Smart" instruments combining real time sensing with tissue manipulation have significant potential to improve surgical practice, especially when combined with real time visualization, robotic devices, and computer-based decision support. In this paper, we have introduced a new class of microsurgical instruments for retinal surgery, incorporating fiber-optic common path Fourier domain OCT sensing. Although the development of such instruments and techniques for using them is still in early stages, our initial experiments are very encouraging.

We have found that very small instruments (~0.5mm diameter) incorporating fiber-optic OCT can be constructed and used to identify tissue boundaries beyond the tip of the instruments. Further, the instrument can be scanned laterally to construct 2D and 3D images from the single A-mode images produced by the basic sensor. We have demonstrated that the sensor can provide real time feedback on the distance from the tool tip to a surface and can be used to enforce safety barriers or support surface following with a robot. Finally, we have shown that these capabilities can be combined to enable user identification of a subsurface target in a scanned image followed by automated placement of the instrument tip on the target. All of these capabilities will be very useful in future clinical vitreoretinal systems.

Further development will require us to address a number of technical and practical questions. The current experiments were performed with the probe perpendicular to the sample surface. Although OCT can identify layers while looking obliquely into tissue at the angles that will be encountered in retinal surgery, it is still necessary to account for approach angle to get correct range data. We are developing several approaches to address this, such as robot pose

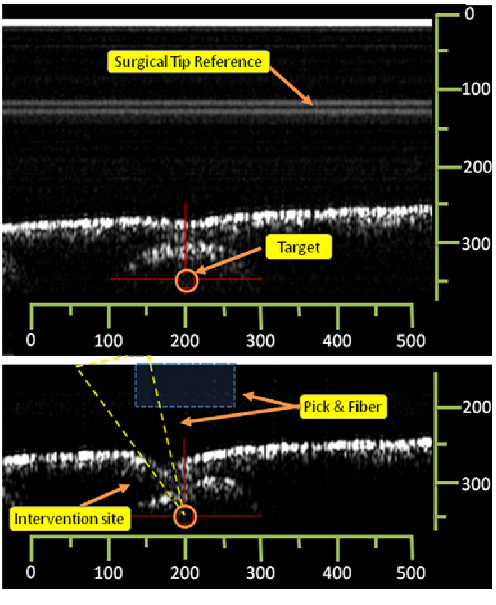


Fig. 6. B-Mode scan before and after intervention

feedback and tracking of tools in the stereo video microscope. Practical issues include: fabrication processes, optical path occlusion by stray particles in the microenvironment, improvement in the speed of OCT data acquisition, and interfacing to our surgical robots. None of these seem insuperable, though experimentation and design iteration will be needed. For example, we need to determine if particle occlusion is a problem within the aqueous environment of the eye, or if the particles simply wash off.

In the immediate future, we will pursue several paths. First, we have just completed a new CP-FDOCT system with considerably higher performance. After repeating the experiments described above, we will begin evaluating the combined system with multiple tissue samples in our lab and will also interface the system to the microsurgical robots in our lab. In the intermediate term, we will also develop methods for tracking instruments in the surgical microscope, for making the necessary depth corrections, for producing B-mode images from the tracked tool positions, and for generating registered overlaid displays on the microscope.

References

1. Sjaarda, R.N., Michels, R.G.: Macular pucker. In: Ryan, S.J. (ed.) Retina, 2nd edn., Mosby, St. Louis, vol. 3, pp. 2301–2312 (1994)
2. Keisuke, M., Gehlbach, P.L., SanoA., D.T., Yoneya, S.: Comparison of Epiretinal Membranes of Differing Pathogenesis Using Optical Coherence Tomography. Retina 24, 57–62 (2004)
3. Fujimoto, J.G., Pitris, C., Boppart, S.A., Brezinski, M.E.: Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy. Neoplasia 2(1-2), 9–25 (2000)
4. Herrmann, J.M., Boppart, S.A., Bouma, B.E., Tearney, G.J., Pitris, C., Brezinski, M.E., Fujimoto, J.G.: Real time imaging of laser intervention with optical coherence tomography. In: Biomedical Optical Spectroscopy and Diagnostics / Therapeutic Laser Applications. OSA Trends in Optics and Photonics Series, paper TSuD2, vol. 22 (1998)
5. Boppart, S.A., Nguyen, F.T., Zysk, A.M., Chaney, E.J., Kotynek, J.G., Oliphant, U.J., Bellafiore, F.J., Rowland, K.M., Johnson, P.A.: Coherent optical imaging and guided interventions in breast cancer: translating technology into clinical applications. In: Proc. SPIE, vol. 6991, 699102 (2008)
6. Han, S., Sarunic, M.V., Wu, J., Humayun, M., Yang, C.: Handheld forward-imaging needle endoscope for ophthalmic optical coherence tomography inspection. J. Biomed. Opt. 13, 20505 (2008)
7. Fleming, I.N., Voros, S., Vagvolgyi, B., Pezzementi, Z., Handa, J., Taylor, R., Hager, G.D.: Intraoperative Visualization of Anatomical Targets in Retinal Surgery. In: Fleming, I.N., Voros, S., Vagvolgyi, B., Pezzementi, Z., Handa, J., Taylor, R., Hager, G.D. (eds.) IEEE Workshop on Applications of Computer Vision, 2008. WAC 2008, January 7-9, 2008, pp. 1–6 (2008)
8. Liu, X., Li, X., Kim, D.-H., Ilev, I., Kang, J.U.: Fiber Optic Fourier-domain Common-path OCT. C. Optics Letters 06(12), 899–903 (2008)
9. Mitchell, B., Koo, J., Iordachita, I., Kazanzides, P., Kapoor, A., Handa, J., Hager, G., Taylor, R.: Development and Application of a New Steady-Hand Manipulator for Retinal Surgery. IEEE ICRA, 623–629 (2007)
10. Riviere, W.A., Khosla, P.: Toward active tremor canceling in handheld microsurgical instruments. IEEE Trans. Rob. Autom. 19, 793–800 (2003)
11. Kapoor, A., Deguet, A., Kazanzides, P.: Software components and frameworks for medical robot control. In: IEEE ICRA, pp. 3813–3818 (2006)

Coronary Tree Extraction Using Motion Layer Separation

Wei Zhang[1], Haibin Ling[2], Simone Prummer[3], Kevin Shaohua Zhou[1], Martin Ostermeier[3], and Dorin Comaniciu[1]

[1] Siemens Corporate Research,755 College Road E. Princeton, NJ 08540, USA
[2] Dept. Computer Science, Temple University, Philadelphia, PA 19122, USA
[3] Siemens AG, Health Care Sector, MED AX PLM-I, Forchheim, Germany

Abstract. Fluoroscopic images contain useful information that is difficult to comprehend due to the collapse of the 3D information into 2D space. Extracting the informative layers and analyzing them separately could significantly improve the task of understanding the image content. Traditional Digital Subtraction Angiography (DSA) is not applicable for coronary angiography because of heart beat and breathing motion. In this work, we propose a layer extraction method for separating transparent motion layers in fluoroscopic image sequences, so that coronary tree can be better visualized.. The method is based on the fact that different anatomical structures possess different motion patterns, *e.g.*, heart is beating fast, while lung is breathing slower. A multiscale implementation is used to further improve the efficiency and accuracy. The proposed approach helps to enhance the visibility of the vessel tree, both visually and quantitatively.

1 Introduction

Coronary angiography is an important enhancement method for the analysis of coronary images and therefore has many clinician applications for cardiac related diseases. However, the task is very challenging because images contain overlaying structures besides blood vessel, *e.g.* Figure 1. Moreover, radiation is kept low for patient's health and thus blood vessels often have poor visibility. *Digital subtraction angiography* (DSA) [7] has been widely used in interventional radiology for enhancing the vessel structure. However, it meets difficulties when applied to the coronary images due to the complicated motion induced by heart beating and breathing.

In this paper we propose a new framework to solve the above problem using motion layer separation. The intuition is based on the fact that different anatomical structure have different motion patterns. For example, lungs usually move in a slow and simple way, heart beating causes much more complicated vessel motion, while bones and spines usually remain static. Consequently, it is natural to model an image sequence as a superposition of different motion layers and then use motion separation to extract layers we are interested in. The basic

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 116–123, 2009.

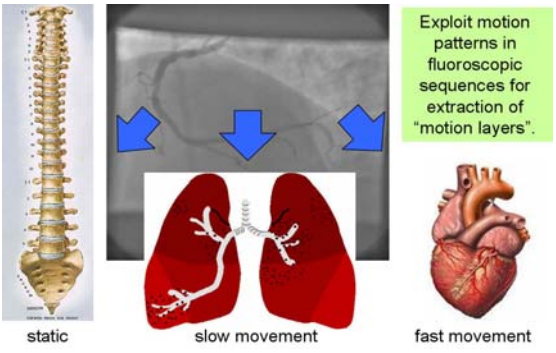


Fig. 1. Different anatomical structures in the scene have different motion patterns

idea is illustrated in Figure 1. The proposed approach is tested on both real and synthetic sequences and demonstrated promising results.

There are several key ingredients in the proposed approach. First, the proposed coronary extraction approach naturally uses motion layer separation. By doing this, it not only avoids the challenging and error prone mask seeking procedure used in traditional DSA approaches, but also achieves robustness by easily integrating neighborhood information. Second, a multiscale scheme is adopted. In addition to achieving better efficiency and accuracy, the multiscale framework naturally maps different layers to different scales. Third, a dense motion field is established by using *thin plate spline* (TPS) [4]. This enables us to handle complicated vessel motions. Fourth, a quantitative measurement consistent to human perception is introduced for evaluating visibility enhancement.

2 Related Work

Traditional motion estimation with adaptation [3] does not model the transparency issue specifically. It does not work for general transparent motion.

Processing in Fourier space. As opposed to the traditional spatial domain approaches, Shizawa and Mase [11] reports that the 3D Fourier transform of a transparent sequence whose transparent layers are in constant translation is made of different planes corresponding to different layers. However, this approach has a major drawback: it assumes a constant transparent motion over a significant number of frames [8].

Explicit transparency modeling. This framework assumes constant motion over three successive frames for two-transparent-layer case. The constraints become cumbersome when we are dealing with more transparent layers. Many classical motion estimation methods have been adapted to the transparency case substituting the brightness assumption with some constraint equations. The popular methods that fall under this category are : block-matching techniques and random fields [12], regularization [13], multi-resolution [1] etc.

Algorithm 1. Multiscale Coronary Extraction

```
/*Motion layer separtion at low resolution of 128×128*/
Data initialization and motion field preparation at low resolution
Solve equation (4) for three layers: background, lung, and vessel
/*Motion layer separtion at middle resolution of 256×256*/
Upsample from low resolution and remove background layer
Data initialization and motion field preparation at middle resolution
Prepare motion field at middle resolution
Solve equation (4) containing two layers for lung, and vessel
/*Final result at high resolution of 512×512*/
Upsample from middle resolution and subtract lung layer at high resolution
```

Parametric models - The Chicken and Egg problem. The detection of multiple motions can be addressed as a segmentation problem. However, the optical flow-field segmentation problem is coupled with the estimation of the flow itself which is a chicken-and-egg problem. If the flow were accurately given everywhere then we can find the motion boundaries. These methods try to solve the segmentation and estimations iteratively. The popular frameworks proposed in the literature are the expectation-maximization framework [14], shift-and-subtract strategy [2], temporal integration to blur the uninterested regions [6].

Repetitive motion. When the dynamics in one layer is assumed to be repetitive, global-to-local space-time alignment can be used to extract the other layer which can have arbitrary motion [9].

3 Our Approach

In this section we describe the proposed approach. We take a hierarchical framework for both efficiency and effectiveness. In this framework, we first subtract the background layer at low resolution. Then the lung and vessel layers are separated at middle resolution. Finally, the vessel layer is refined at the finest scale. A summary of the approach is given in Algorithm 1.

3.1 Problem Formulation

Without loss of generality, we assume that our problem is to extract layers from image sequences which consist of N layers overlapping with uniform transparency, i.e., the contribution of each layer is $1/N$. For X-ray image formation, an exponential relationship exists between the incident and transmitted photon fluence [10]. After passing through multiple layers of materials, the final amount of photon fluence received by the detector which forms the X-ray image is:

$$N_x = N_0 e^{\sum \mu_i x_i}, \tag{1}$$

where μ_i is attenuation coefficient of the i_{th} layer and x_i is its thickness.

Because of the exponential form of the X-Ray image formation, X-ray images are usually manipulated in the logarithmic space. In the logarithmic image space, the image can be written as a linear combination of the layers:

$$I_m = \sum_{l=1}^{N} T_l^m L_l, \tag{2}$$

where I_m is the m^{th} observed image, L_l is the l^{th} unknown layer. T_l^m is the motion field for the l^{th} layer that maps L_l to I_m. Refer to Sec. 3.3 for how the motion field is constructed. I_m and L_l have the same size. Denote W and H as their width and height, respectively. L_l has $W \times H$ pixels with unknown values. Our goal is to find images of different layers by minimizing the reconstruction error. For achieving robustness to noise, we use $M(M > N)$ images to find a least square solution to the following equation

$$\min \sum_{m=1}^{M} \| \sum_{l=1}^{N} T_l^m L_l - I_m \|^2 \quad . \tag{3}$$

To solve (3), it is transformed into a least square optimization form. We omit the details because of the page limit.

$$\arg\min_x \|Cx - d\|^2 \ , \tag{4}$$

where $C \in \mathbb{R}^{K \times J}$ is a matrix that is constructed based on the motion field, $x \in \mathbb{R}^J$ is a vector that is the concatenation of all the unknown pixel values of all N layers, $d \in \mathbb{R}^K$ is a vector which is the concatenation of all the (known) pixel values of all the M images. $K = W \times H \times M$ and $J = W \times H \times N$.

In practice, the number of images M should be relatively big to tolerate small shape deformation in one image. On the other hand, it should not be too big because shape deformation would accumulate with more images, which would violate the assumption that sum of all layers equal to the observed image. In our experiments, we empirically found that $M = 9$ produces the best result. This coincides with the fact that the number of frames per heart cycle is roughly 9. The workflow is the following: for each image I_m in the sequence, we use images in a local time window $I_{m-4}, I_{m-3}, \ldots, I_{m+4}$ and the corresponding motion fields to construct C and d in Equation 4. Once the optimization converges to a solution x, we re-arrange x and obtain all the transparent layers in that image. After all the images have been processed, a video sequence is obtained for each layer.

3.2 Multiscale Framework

The least squares problem (4) is very large in scale. For example, to estimate 3 layers for a sequence with image size 512×512, the number of unknowns is 786432. On the other hand, the matrix C is very sparse with most of its entries to be zero. So we can resort to the iterative optimization technique to solve the problem. Notice that the value of unknowns can not be negative, we are

actually trying to solve a constrained optimization problem. In the current implementation, the solution is found using a trust region method based on the interior-reflective Newton method. In each iteration, an approximate solution is obtained using the method of preconditioned conjugate gradients.

For the coronary angiography, we could assume that there are 3 layers in the image: the static background like rib, the slow moving lung and fast beating vessels. To reduce the complexity of the problem, we perform the estimation in a pyramid. For the image of size 512×512, we build additional two levels of size 256×256 and 128×128. Assuming that the background remain unchanged through out the sequence, we estimate it first using the lowest resolution. Following the workflow mentioned above, we could get a sequence of estimated background images. Most of them are quite similar but not necessarily the same. There might be few erroneous estimates because of the correlation in motions of different layers. Nevertheless, the median of of all those estimation gives a reliable estimate of the overall background. Then the estimated background is upsampled to 256×256 and we subtract the background from each image for estimating the remaining two moving layers (vessel tree and diaphragm). As Figure2 illustrated, only the moving objects are retained after subtracting the background and they look more evident than in original image.

Now the new (foreground) sequence contains only two transparent layers. We follow the same procedure to estimate the diaphragm layer and vessel layer in the middle resolution (256×256). After reducing the number of layers, the number of unknowns in Equation 4 is now $256 \times 256 \times 2 = 131072$. Then we upsample the estimated diaphragm layer to 512×512 and subtract it from the foreground sequence, thus obtain the final vessel layer at the original resolution. By doing the whole process in a pyramid, we keep pushing details into the vessel layer because this is really what the clinician want. In addition, we avoid the big problem of trying to optimize 786432 unknowns all together. This not only brings the huge advantage in efficiency, but also helps to avoid some local maxima for the direct optimization as mentioned above. The idea is to takes advantage of the fact that the background layer is invariant across time and slow moving layer is relatively invariant compared to vessel, so that a large part of the search space can be eliminated.

3.3 Constructing Motion Field

The motion of vessels is very complicated, global transformation like affine transformation [1] is therefore insufficient to model it. Instead, we use a dense motion field to represent the vessel motion. Specifically, for any location with coordinates (x, y) at layer L_l, the motion field at this particular coordinates $T_l^m(x, y)$ maps (x, y) to a new position (x', y') in image I_m.

An efficient way to construct the motion field is through the *thin plate spline* (TPS) [4] interpolation. Given two point sets with correspondence between them, TPS finds a nonlinear warping by minimizing a second order "bending energy". In our task, the TSP warping has the following formula

$$\begin{cases} x' = a_x x + b_x y + c_x + \sum_{j=1}^{n} d_{x,j} U((x,y) - (x_j, y_j)) \\ y' = a_y x + b_y y + c_y + \sum_{j=1}^{n} d_{y,j} U((x,y) - (x_j, y_j)) \end{cases}, \tag{5}$$

where $\{(x_j, y_j)\}_{j=1}^{n}$ is the sparse anchoring point set; $a_i, b_i, c_i, d_{i,j}$ for $i = \{x, y\}$, $j = 1..n$ are warping parameters that are estimated from sparse motion vectors; and $U(.)$ is a Radial Basis Function. We use manually selected control points (e.g. junction points on vessels) to get sparse motion vectors since this work focuses on motion layer separation. For automatic estimation of sparse motion vectors, We have developed techniques for automatic detection [15] and tracking [16] of thin curved structures, which could be used for vessel detection and tracking in low resolution images. The detection of junction points can be done using the Marginal Space Learning Framework [17].

4 Experimental Results

4.1 Experiments on Real Sequences

We tested the proposed approach on two real sequences, one for vessel enhancement and another for guidewire enhancement. Figure 2 shows the extracted three layers for one image. As can be seen, static objects like bones are retained in the background layer. Slow moving diaphragm is extracted in another layer. The vessel tree we are interested in lies in its exclusive layer and is more visible than that in the original fluoroscopic image. We also applied the method to separating the guidewire and lung in another sequence and the result is shown in Figure 3. Please refer to supplemental videos for more results.

4.2 Evaluation of Visibility Enhancement

While we can see the visibility improvement qualitatively, it's desired to quantitatively analyze the enhancement. A synthetic sequence is shown in Figure 4(a).

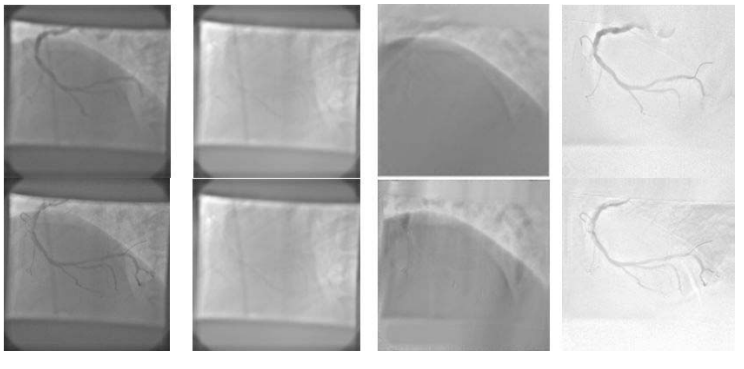


Fig. 2. Vessel tree separation. From left to right: input fluoroscopic image, background, diaphragm, and vessel tree.

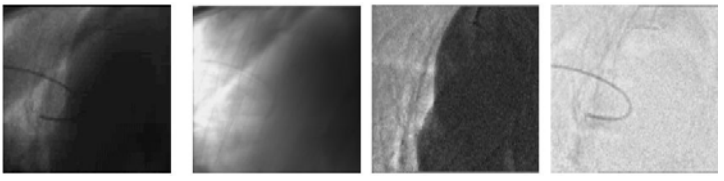
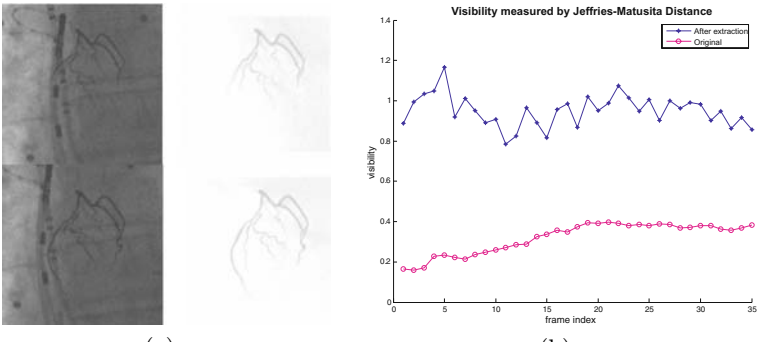


Fig. 3. Guidewire separation. From left to right: original fluoroscopic image, background, lung, and guidewire. The guide wire is hard to see before separation.



(a) (b)

Fig. 4. (a) Synthetic image with moving background (left), extraction result (right). (b) Quantitative evaluation of visibility enhancement.

The measurement we use is based on the Jeffries-Matusita distance, which measures the separability between two classes, vessel and background. The higher the measure, the better the separability.

$$J_{vb} = 2(1 - e^{-B_{vb}}), \qquad B_{vb} = \frac{(\mu_v - \mu_b)^2}{4(\sigma_v^2 + \sigma_b^2 + \epsilon)} + \frac{1}{2} \ln \frac{\sigma_v^2 + \sigma_b^2}{2\sigma_v \sigma_b}, \tag{6}$$

where μ_v and μ_b are means of vessel and background, respectively and σ is the standard deviation. Note there is a term ϵ in the the formulation, which is added to to make it consistent with the human perception. We have experimentally set $\epsilon = 100$ according to the input of multiple experts. For the synthetic image in Figure 4(a), $J_{vb} = 0.31$, while the extracted result in the right has $J_{vb} = 0.89$. The evaluation of the whole sequence is summarized in Figure 4(b). We have carried out extensive experiments with different noise adding to the synthetic image. They all exhibited clearly quantitatively enhancement.

5 Conclusion

In this paper, we propose a framework for separating transparent layers from fluoroscopic image sequences, so that coronary tree can be better visualized. It

utilizes motion information to decompose an image into different layers. The proposed approach does not require any pre-selected mask, thus avoid the difficulties of the traditional DSA. The motion layer extraction is proceeded in a hierarchical fashion to achieve both efficiency and robustness. The proposed approach is tested on both real and synthetic sequences, and promising results are observed. Our ongoing work focus on automatic motion vector detection and learning-based tracking of vessel structures.

References

1. Auvray, V., Liénard, J., Bouthemy, P.: Multiresolution parametric estimation of transparent motions and denoising of fluoroscopic images. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 352–360. Springer, Heidelberg (2005)
2. Bergen, J.R., Burt, P.J., Hingorani, R., Peleg, W.: A three-frame algorithm for estimating two component image motion. IEEE PAMI 14(9), 886–895 (1992)
3. Black, M., Anandan, P.: The robust estimation of multiple mo- tions:parametric and piecewise-smooth fields. CVIU 63(1), 75–104 (1996)
4. Bookstein, F.: Principal Warps: Thin-Plate-Splines and Decomposition of Deformations. IEEE Trans. PAMI 11(6), 567–585 (1989)
5. Irani, M., Peleg, S.: Motion analysis for image enhancement: Resolution, occlusion, and transparency (1993)
6. Irani, M., Rousso, B., Peleg, S.: Computing occluding and transparent motions. IJCV 12(1), 5–16 (1994)
7. Meijering, E., Niessen, W., Viergever, M.: Retrospective Motion Correction in Digital Subtraction Angiography: A Review. IEEE. Med. Imaging 18(18) (1999)
8. Pingault, M., Pellerin, D.: Motion estimation of transparent objects in the frequency domain. J. of Signal Processing 84(4), 709–719 (2004)
9. Sarel, B., Irani, M.: Separating transparent layers through layer information exchange. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3024, pp. 328–341. Springer, Heidelberg (2004)
10. Seibert, J.A., Boone, J.M.: X-ray imaging physics for nuclear medicine technologists 2: X-ray interactions and image formation. J. Nuclear Med. Tech. (2005)
11. Shizawa, M., Mase, K.: Principle of superposition: A common computational framework for analysis of multiple motion, pp. 164–172 (1991)
12. Stuke, I., Aach, T., Barth, E., Mota, C.: Estimation of multiple motions using block-matching and markov random fields (2004)
13. Toro, J., Ownes, F., Medina, R.: Multiple motion estimation and segmenta- tion in transparency. In: ICASSP (2000)
14. Weiss, Y., Adelson, E.: A unified mixture framework for motion segmentation: Incorporating spatial coherence and estimating the number of models. In: CVPR (1996)
15. Barbu, A., Athitsos, V., Georgescu, B., Comaniciu, D.: Hierarchical Learning of Curves: Application to Guidewire Localization in Fluoroscopy. In: CVPR (2007)
16. Wang, P., Zhu, Y., Chen, T., Zhang, W., Zhou, S., Comaniciu, D.: Robust Guidewire Tracking in Fluoroscopy. In: CVPR (2009)
17. Zheng, Y., Barbu, A., Comaniciu, D.: Fast Automatic Heart Chamber Segmentation from 3D CT Data Using Marginal Space Learning. In: ICCV (2007)
18. Nakaura, T., Ooishi, S., Tayama, S., Ogawa, H.: A New Technique for Coronary Digital Subtraction Angiography. Scientific Meeting Japan. Circul. Society (2006)

A Fast Alternative to Computational Fluid Dynamics for High Quality Imaging of Blood Flow

Robert H.P. McGregor[1], Dominik Szczerba[1,2], Krishnamurthy Muralidhar[3], and Gábor Székely[1]

[1] Computer Vision Laboratory, Sternwartstr. 7, 8092 Zürich, Switzerland
mcgregor@vision.ee.ethz.ch
[2] IT'IS Foundation for Research, Zürich, Switzerland
[3] Department of Mechanical Engineering, Indian Institute of Technology Kanpur, Kanpur 208 016, UP, India

Abstract. Obtaining detailed, patient-specific blood flow information would be very useful in detecting and monitoring cardio-vascular diseases. Current approaches rely on computational fluid dynamics to achieve this; however, these are hardly usable in the daily clinical routine due to the required technical supervision and long computing times. We propose a fast measurement enhancement method that requires neither supervision nor long computation and it is the objective of this paper to evaluate its performance as compared to the state-of-the-art. To this end a large set of abdominal aortic bifurcation geometries was used to test this technique and the results were compared to measurements and numerical simulations. We find that this method is able to dramatically improve the quality of the measurement information, in particular the flow-derived quantities such as wall shear stress. Additionally, good estimation of unmeasurable quantities such as pressure can be provided. We demonstrate that this approach is a practical and clinically feasible alternative to fully-blown, time-consuming, patient-specific flow simulations.

1 Introduction

High quality imaging of blood flow patterns can provide useful information for understanding cardio-vascular pathologies, predicting their onset and choosing an optimal therapy strategy. There is particular interest in measuring wall shear stress (WSS) as this plays an important role in the development of atherosclerosis and other flow-related diseases such as aneurysms. Unfortunately WSS is extremely difficult to measure as it is related to the derivative of velocity close to the wall, precisely where most conventional imaging techniques are least reliable. The use of both phase contrast magnetic resonance imaging (PC-MRI) and Doppler ultrasound, which are able to resolve flow fields in vivo and non-invasively in complex geometries have been studied extensively (e.g. [1] or [2]). Although they are found to be well suited for estimating the approximate time

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 124–131, 2009.

and spatially-averaged WSS patterns, both these modalities suffer from poor resolution and, even more importantly, from inadequate accuracy near the wall due to partial volume effects or wall motion induced artifacts, thus making them unable to provide high quality WSS estimates. However, these limitations can effectively be overcome by using computational fluid dynamics (CFD) in conjunction with medical imaging [3], which allows for very detailed flow field representations. Moreover, CFD is able to provide any desired (and usually non-measurable) flow quantities such as vorticity or pressure distributions which can be very useful for, e.g., aneurysm rupture prediction. The richness of the CFD simulation data has motivated a large research effort to investigate its use in patient-specific flow imaging (see [4] for a review) and there is much hope that these results could significantly contribute to the improvement of the diagnosis and treatment of vascular diseases such as cerebral aneurysms [5]. However the use of computer simulations requires in-depth numerical knowledge and enormous computational power and time, which makes them impractical for routine patient care and has kept them out of the clinics for the time being.

We have already proposed a clinically feasible approach to the combination of patient-specific measurement data and CFD [6]. This previous work focussed on exploring the potential of such an enhancement technique on a simplified 2D case. We now evaluate the quality of such 3D flow field reconstructions using a large database of abdominal aortic bifurcation geometries. We focus in particular on the quality of the WSS estimations, as this type of secondary flow information is extremely sensitive to measurement accuracy. Acquisition of several hundred real patient cases is a dedicated effort in itself and is not the focus of our current study. We therefore use artificially generated but statistically meaningful geometries. This approach conveniently provides us with a ground truth to evaluate our results and simultaneously reduces artifacts caused by faulty reconstruction of the geometry. In practice, this measurement enhancement technique could be applied to any imaging modality which is able to detect the geometry and preferably some flow velocity data, but we have chosen to focus on PC-MRI as this can measure both geometry and time-resolved flow velocity fields non-invasively.

2 Methods

Enhancement method. The method is fully described in our previous work [6] and will just shortly be summarized here for completeness. We use a large database of pre-computed flow fields for a given anatomical location and use these to perform an interpolation of actually measured, sparse and noisy velocity data, allowing us to recover a patient-specific, high quality flow field. Principal component analysis is applied to the database and the enhancement is performed as an optimisation in frequency space, using the method proposed by Blanz and Vetter [7], analogously to the procedure used in statistical shape models. For a basic overview of this enhancement method see figure 1. The main difference to our previous 2D implementation lies in the mesh generation, mesh mapping and its associated coordinate transformation. Indeed, to achieve this in 3D with

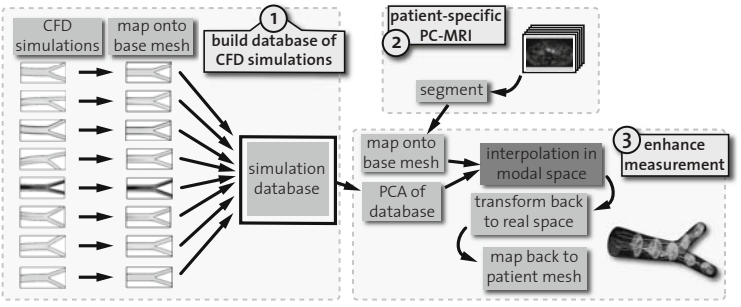


Fig. 1. Method flowchart

bifurcation geometries a radial coordinate system was used, the axial component being recovered by solving the Stokes equations with constant pressure boundary conditions at the inlet ($p1 = 1$) and both outlets ($p2 = 0$). In this way a coordinate system common to all bifurcations could be generated, thus making the velocity components comparable across geometries.

Building the database. As mentioned earlier, the database used here is built entirely from artificial data for evaluation purposes. A base mesh having the approximate physiological dimensions of an abdominal aortic bifurcation and being symmetric with respect to the sagittal plane was first generated. This was then deformed using 18 geometric parameters which were randomly varied within healthy physiological variability as described by Sun et al. [8] to create the various meshes. Additionally a random deformation field was applied to each geometry thus perturbing the vessel wall with high frequency local variations. Fifteen randomly selected bifurcations from the database can be seen in figure 4(f). Once the meshes became available, oscillatory flow simulations were performed in each one, using velocity boundary conditions at the inlet and fixed pressure at the outlet. Skewed parabolic velocity fields were used at the inlet, randomly varied with respect to oscillatory frequency, skewness and amplitude to mimic a variety of upstream conditions. The pressure was set to 0 at both outlets, using the assumption that the downstream conditions are the same for both iliac branches, which is physiologically reasonable. Each simulation was run for 3 cardiac cycles so as to reduce transient effects and the results from the last cycle were taken as our solutions. A total of 600 flow fields were generated in this fashion, each having its own specific geometry and boundary conditions.

Recovering the pressure fields. In addition to the velocities, the pressure fields were also stored in the database. This allows us to recover the patient-specific pressure distribution at very little extra computational cost, which can be interesting especially in the case of aneurysm rupture prediction, where the local pressure field plays an important role.

Evaluating the accuracy of the method. In order to evaluate our method, 100 of the flow fields were randomly selected for leave-one-out tests and therefore removed from the database. We then simulated PC-MRI measurements of each of these test cases. This was done in two steps. First we defined a typical voxel

size of $1 \times 1 \times 4\,mm$ and averaged the velocities inside each voxel for 20 equally-spaced timesteps. Secondly, we added white Gaussian noise with a maximal intensity of 5% of the velocity encoding value (VENC). At voxels which are on the boundary, the partial volume effect typical to PC-MRI was simulated by adding nodes with 0 value outside the geometry before averaging. We then compared the velocity, WSS and pressure distributions obtained from both the simulated flow measurements and from our enhancement method with the ground truth, given by the original simulations. WSS was computed in all cases by using quadratic shape functions of the tetrahedra having a face on the wall, so the acquisition data had first to be interpolated onto the mesh, similarly to Cheng et al. [9].

3 Results

Having applied the enhancement for 100 test cases, the most obvious way to present the results is a statistical evaluation: this can be found in table 1. For each quantity of interest and for each case we performed a linear regression analysis of the type seen in figures 2 and 3 to evaluate how well the measured and the enhanced [1] values compared to the simulated ones. We then considered the r^2 correlations (the square of Pearson's correlation coefficient) and the slope of the fitted linear functions as these best characterize the fit. The closer both these numbers are to 1, the better the observation matches the known data. In addition to this statistical evaluation we show the simulated, measured and enhanced flow, WSS and pressure fields for a representative case in figure 4.

Table 1. Mean values and standard deviations over the 100 test cases of the r^2 correlation coefficients and the slopes of the fitted linear curves relating both the measured and the enhanced data to the simulations (ground truth) for all components of velocity (u, v, w), the norm of velocity $(|U|)$, the WSS and the pressure

	Measured		Enhanced	
	r^2	slope	r^2	slope
u	0.873 / 0.024	0.818 / 0.028	0.896 / 0.058	0.897 / 0.051
v	0.897 / 0.017	0.834 / 0.022	0.917 / 0.053	0.915 / 0.050
w	0.883 / 0.023	0.876 / 0.018	0.980 / 0.012	0.971 / 0.014
$\|U\|$	0.842 / 0.029	0.871 / 0.020	0.976 / 0.013	0.974 / 0.013
WSS	0.824 / 0.023	0.359 / 0.023	0.947 / 0.059	0.929 / 0.048
$pressure$			0.996 / 0.007	0.938 / 0.131

Velocity. Figures 4(a), 4(b) and 4(c) show typical flow fields during end diastole. This phase of the heart cycle was chosen as the velocities are low and recirculations near the boundary take place. In all figures a detail of the streamlines

[1] We will systematically refer the original (gold standard) simulation data as *simulated*, to the simulated PC-MRI data as *measured* and to the measurement data after being enhanced by our method as *enhanced*.

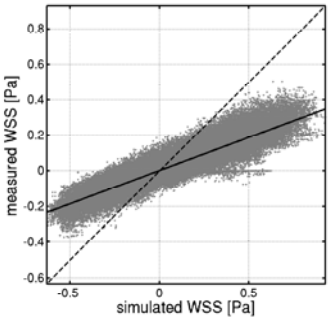


Fig. 2. Correlation of the measured with the simulated WSS

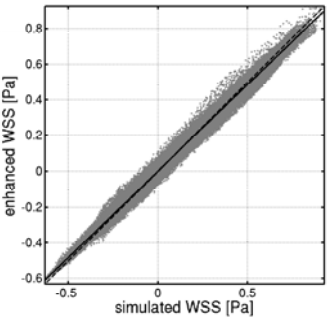


Fig. 3. Correlation of the enhanced with the simulated WSS

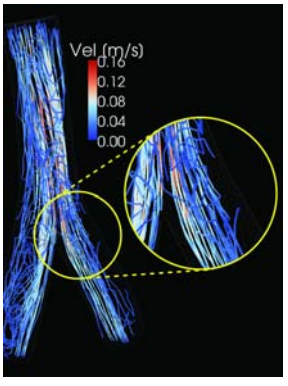
is shown, demonstrating that these recirculations are not resolved by the measurement, but do appear correctly in the enhanced flow field. However, despite being sparser and noisier, the measured flow fields are still reasonably accurate, as shown by the statistical data in table 1.

WSS. Measured, enhanced and simulated WSS distributions at peak systole and end diastole are shown in figures 4(d) and 4(e) respectively. One may observe that the WSS as postprocessed from the measurement is clearly underestimated. However, its spatial distribution, which is difficult to perceive directly from this illustration, is approximately correct. Figures 2 and 3 show the statistical correlations between the measured and simulated and between the enhanced and simulated WSS fields respectively for a representative case. It can be seen by the slopes of the regression plots that the absolute values of the measured WSS are significantly smaller than the simulated ones, whereas the enhanced ones are in the correct range. This is mostly due to the partial volume effect which leads to underestimation of the velocities close to the wall. The results were similar in all cases, as shown by the poor slope value for the measured WSS in table 1. The high correlation and slope values obtained by the enhancement method indicate that it is able to reconstruct WSS patterns with very good accuracy and could thus effectively replace full CFD simulations for this task.

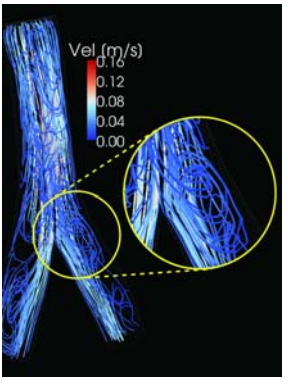
Pressure. We show no data for the measured pressure fields, as there is no way of measuring a full pressure field by any known in-vivo, non-invasive method. Despite having no measured pressure information, the reconstructed pressure fields were very well correlated with the simulated ones (see table 1) and the close similarity of the pressure distributions at peak systole and end diastole shown in figures 4(g) and 4(h) is typical for all cases.

4 Discussion

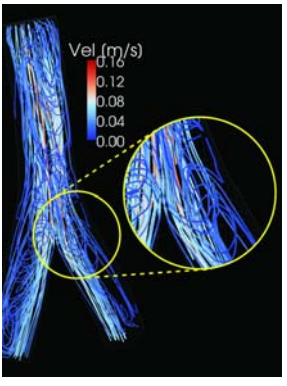
The fact that both the measured and enhanced velocities correlate well with the simulated ones confirms the work of other authors (see e.g. [10] or [11])



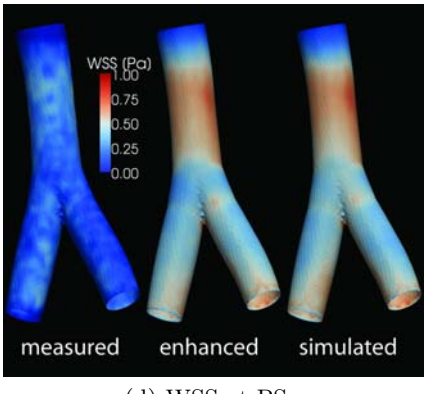
(a) Measured velocity.



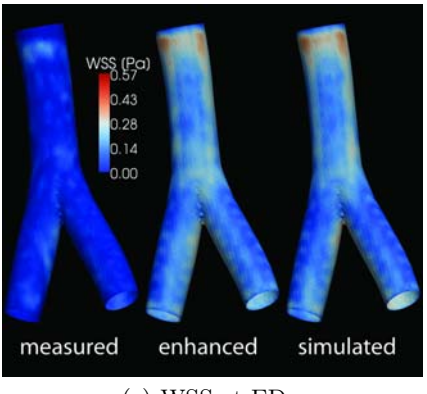
(b) Enhanced velocity.



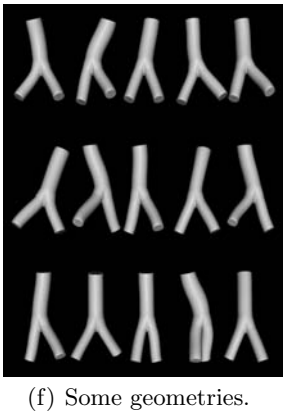
(c) Simulated velocity.



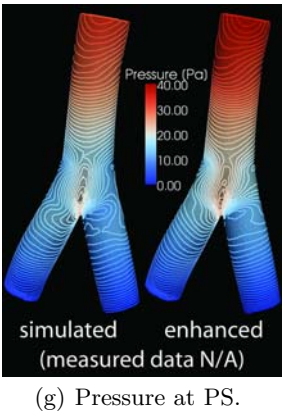
(d) WSS at PS.



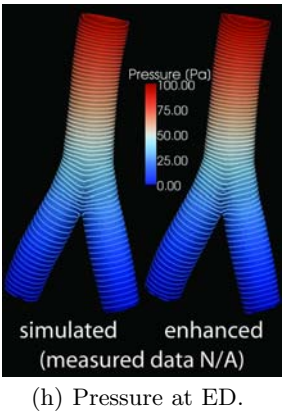
(e) WSS at ED.



(f) Some geometries.



(g) Pressure at PS.



(h) Pressure at ED.

Fig. 4. Streamlines of velocity (top row) at end diastole, with zoom on a recirculation, WSS (middle row) and pressure (bottom row) at both peak systole (PS) and end diastole (ED) and 15 geometries from the database (bottom left)

who also conclude that PC-MRI is adequate for acquiring the velocity field. However it must be remembered that the enhanced field can have any desired spatio-temporal resolution, whereas the measured PC-MRI data is limited in this respect. Also we focused here on the case of the abdominal aortic bifurcation which is a relatively large vessel. In smaller vessels the resolution limits of this modality will very quickly become prohibitive. However the inability to resolve lower velocities and those near the endothelium has a disastrous effect on the WSS predictions. As shown here and confirmed by others, [12] the WSS values as obtained directly from PC-MRI are by far not accurate enough, in particular with respect to their amplitudes. The approximate location of regions of low or high WSS can be adequately obtained, but only time- and spatially-averaged values can be considered. Should one desire a better quality WSS prediction, information from CFD must be used and our method is able to provide this with high accuracy, without needing the enormous computational resources required for such simulations.

We focused here on two flow quantities other than velocity: WSS and pressure, but obviously any other flow-related variable such as vorticity or oscillatory shear index can also be addressed by this enhancement strategy. Another interesting extension would be the integration of full fluid-structure interaction (FSI). In this study we considered fixed walls, but adding FSI is possible in a straightforward manner. It would require much more computational effort to generate the database but would have little or no impact on the speed of the enhancement procedure while delivering a lot more information such as intramural wall stress.

We focused on healthy, i.e. relatively regular geometries which, while still showing very complex flow patterns, have much less variability than pathological cases such as aneurysms. However, addressing these will not require any modifications of the overall framework, only that the database be adequately larger to account for the added variability.

5 Conclusions and Outlook

We have shown a 3D extension of our previously presented flow measurement enhancement strategy and evaluated its performance using a large, representative, synthetic dataset. We were able to demonstrate that it provides reliable predicions about flow and clinically important flow-related quantities, clearly surpassing the quality of direct measurements. We conclude that this is an attractive alternative to costly, fully-blown CFD simulations as it is able to achieve similar quality while being orders of magnitude faster (minutes rather than hours or even days). Future work in this direction will now focus on using real patient data and pathological variability, both to build the database and to test the enhancement. However, this evaluation has shown that our approach is a significant step towards bringing the full richness of CFD into daily clinical practice.

Acknowledgments. This work was supported by the Indo-Swiss Joint Research Programme.

References

1. Meckel, S., Stalder, A., Santini, F., Radü, E.W., Rüfenacht, D., Markl, M., Wetzel, S.: In: vivo visualization and analysis of 3-D hemodynamics in cerebral aneurysms with flow-sensitized 4-D MR imaging at 3 T. Neuroradiology 50(6), 473–484 (2008)
2. Tortoli, P., Morganti, T., Bambi, G., Palombo, C., Ramnarine, K.V.: Noninvasive simultaneous assessment of wall shear rate and wall distension in carotid arteries. Ultrasound in Medicine & Biology 32(11), 1661–1670 (2006)
3. Papathanasopoulou, P., Zhao, S., Köhler, U., Robertson, M.B., Long, Q., Hoskins, P., Xu, X.Y., Marshall, I.: MRI measurement of time-resolved wall shear stress vectors in a carotid bifurcation model, and comparison with CFD predictions. Journal of Magnetic Resonance Imaging 17(2), 153–162 (2003)
4. Steinman, D.A., Vorp, D.A., Ethier, C.R.: Computational modeling of arterial biomechanics: Insights into pathogenesis and treatment of vascular disease. Journal of Vascular Surgery 37(5), 1118–1128 (2003)
5. Sforza, D.M., Putman, C.M., Cebral, J.R.: Hemodynamics of cerebral aneurysms. Annu. Rev. Fluid Mech. 41(1), 91–107 (2009)
6. McGregor, R.H.P., Szczerba, D., von Siebenthal, M., Muralidhar, K., Székely, G.: Exploring the use of proper orthogonal decomposition for enhancing blood flow images via computational fluid dynamics. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 782–789. Springer, Heidelberg (2008)
7. Blanz, V., Vetter, T.: Reconstructing the complete 3D shape of faces from partial information. Informationstechnik und Technische Informatik 44(6), 295–302 (2002)
8. Sun, H., Kuban, B., Schmalbrock, P., Friedman, M.: Measurement of the geometric parameters of the aortic bifurcation from magnetic resonance images. Annals of Biomedical Engineering 22(3), 229–239 (1994)
9. Cheng, C.P., Parker, D., Taylor, C.A.: Quantification of wall shear stress in large blood vessels using lagrangian interpolation functions with cine phase-contrast magnetic resonance imaging. Annals of Biomedical Engineering 30(8), 1020–1032 (2002)
10. Boussel, L., Rayz, V., Martin, A., Acevedo-Bolton, G., Lawton, M.T., Higashida, R., Smith, W.S., Young, W.L., Saloner, D.: Phase-contrast magnetic resonance imaging measurements in intracranial aneurysms in vivo of flow patterns, velocity fields, and wall shear stress: Comparison with computational fluid dynamics. Magnetic Resonance in Medicine 61(2), 409–417 (2009)
11. Hollnagel, D.I., Summers, P.E., Kollias, S.S., Poulikakos, D.: Laser doppler velocimetry (LDV) and 3D phase-contrast magnetic resonance angiography (PC-MRA) velocity measurements: Validation in an anatomically accurate cerebral artery aneurysm model with steady flow. Journal of Magnetic Resonance Imaging 26(6), 1493–1505 (2007)
12. Katritsis, D., Kaiktsis, L., Chaniotis, A., Pantos, J., Efstathopoulos, E.P., Marmarelis, V.: Wall shear stress: Theoretical considerations and methods of measurement. Progress in Cardiovascular Diseases 49(5), 307–329 (2007)

Interventional 4-D Motion Estimation and Reconstruction of Cardiac Vasculature without Motion Periodicity Assumption*

Christopher Rohkohl[1,2], Günter Lauritsch[2], Marcus Prümmer[1], and Joachim Hornegger[1]

[1] Chair of Pattern Recognition, Department of Computer Science, Friedrich-Alexander University Erlangen-Nuremberg, 91058 Erlangen, Germany
[2] Siemens AG, Healthcare Sector, 91301 Forchheim, Germany

Abstract. Anatomical and functional information of cardiac vasculature is a key component of future developments in the field of interventional cardiology. With the technology of C-arm CT it is possible to reconstruct intraprocedural 3-D images from angiographic projection data. Current approaches attempt to add the temporal dimension (4-D) by ECG-gating in order to distinct physical states of the heart. This model assumes that the heart motion is periodic. However, frequently arrhytmic heart signals are observed in a clinical environment. In addition breathing motion can still occur. We present a reconstruction method based on a 4-D time-continuous motion field which is parameterized by the acquisition time and not the quasi-periodic heart phase. The output of our method is twofold. It provides a motion compensated 3-D reconstruction (anatomic information) and a motion field (functional information). In a physical phantom experiment a vessel of size 3.08 mm undergoing a non-periodic motion was reconstructed. The resulting diameters were 3.42 mm and 1.85 mm assuming non-periodic and periodic motion, respectively. Further, for two clinical cases (coronary arteries and coronary sinus) it is demonstrated that the presented algorithm outperforms periodic approaches and is able to handle realistic irregular heart motion.

1 Introduction

One key component of future developments in the field of interventional cardiology is three-dimensional image information before, during and after interventional procedures, e.g. pacemaker implantations or ablation procedures. With the technology of C-arm CT it is possible to reconstruct intraprocedural 3-D images from angiographic projection data [1]. However, cardiac reconstruction is yet a challenging problem due to the long acquisition time of several seconds at which a couple of heart beats occur, leading to motion related image artifacts, e.g. blurring or streaks.

* We thank Prof. Dr. med. J. Brachmann, Dr. med. K. Gutleben and Dr. med. H. Rittger from the Klinikum Coburg, Germany for providing clinical image data.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 132–139, 2009.

An established technique for time-resolved cardiac reconstruction is to record the electrocardiogram (ECG) during the data acquisition. Based on the ECG-signal a relative cardiac phase is assigned to each projection image assuming a cyclic heart motion [2]. The phase information is used for a phase-correlated reconstruction by gating or motion estimation and compensation. A gated reconstruction takes only those images into account that lie inside a defined temporal window, that is centered at the targeting reconstruction phase [3,4]. This is however not ideal in terms of missing data and residual motion. To increase the data usage motion compensated reconstruction algorithms are applied. The phase information is used during motion estimation to parameterize a motion field that maps every heart phase to the target phase by some kind of registration operation [5,6,7,8]. The common problem of such approaches is that the averaged periodic motion model does not necessarily represent accurately the actual heart motion of each individual beat. Thus the quality of the motion correction and periodicity assumption are correlated.

Accordingly, the previous methods were shown to provide reasonable results in the presence of regular heart rates without breathing or other patient motion. However, in the field of intraprocedural cardiac reconstruction, the patients suffer from heart diseases and cannot completely hold breath, stay still or have irregular heart beats. Those aspects do conflict with the periodicity assumption. Up to now, these problems were addressed by approximate 2-D corrections in the projection image. Blondel et al. [5] proposed to model breathing motion of the heart as a translation mainly in axial direction. Hansis et al. [9] proposed to cope with the problem by performing a 2-D/2-D registration of the projection image with a forward projection of an initial reconstruction. However, none of the methods can cope with the general case of non-cyclic 3-D motion. In this paper a method for the 4-D reconstruction of selectively contrasted cardiac vasculature without perodicity assumption is introduced. This is achieved by two major contributions: First, the introduction of a time-continous motion field which is parameterized by the acquisition time and not the periodic heart phase. Second, the development of a theoretically founded, fast and robust motion estimation algorithm which is based on an extension of a FDK reconstruction to dynamic objects and thus couples motion estimation and reconstruction.

2 Methods

The presented motion estimation scheme is built up from two components. First, an initial reference reconstruction is performed using an ECG-gated or even non-gated reconstruction. This reference image is post-processed such that only the vasculature is visible. Second, the parameters of the motion model are estimated by optimizing an objective function. The objective function aims to maximize the joint intensity of the reference image and the motion compensated FDK reconstruction.

This part of the paper is organized to contain a step-wise development of the required methods. In Sect. 2.1 the parametric motion model is presented.

Next, the motion model is introduced into the FDK reconstruction in Sect. 2.2. Prepared with both concepts, the objective function for motion estimation is subject of Sect. 2.3. Finally, in Sect. 2.4 the optimization strategy is discussed.

2.1 4-D B-Spline Motion Model

We assume a time-continous motion model that maps a voxel $\boldsymbol{x} = (x_0, x_1, x_2)^T$ to a new voxel location $\boldsymbol{x}'$ for each projection image. It is conceptionalized by a function $M : \mathbb{N} \times \mathbb{R}^3 \times \mathsf{P} \mapsto \mathbb{R}^3$ with $M(i, \boldsymbol{x}, \boldsymbol{r}) = \boldsymbol{x}'$ transforming the voxel coordinate $\boldsymbol{x}$ at the time of the i-th projection image. The mapping is based on the motion model paramters $\boldsymbol{r} \in \mathsf{P}$. In this work, a 4-D B-spline is used. A set of $C_j \times C_k \times C_l \times C_t$ control points is placed uniformly in space and time. Each control point is assigned a displacement vector, forming the set of motion model parameters $\mathsf{P} = \{\boldsymbol{r}_{jklt} \in \mathbb{R}^3 \,|\, 1 \leq j, k, l, t \leq C_j, C_k, C_l, C_t\}$. Formally, the motion model is then given by

$$M(i, \boldsymbol{x}, \boldsymbol{r}) = \boldsymbol{x} + \sum_{j,k,l,t} B_j(x_0) B_k(x_1) B_l(x_2) B_t(i)\, \boldsymbol{r}_{jklt} \quad \forall j, k, l, t\,, \tag{1}$$

with $\{B.\}$ being the set of cubic B-spline basis functions [10].

2.2 Dynamic FDK-Reconstruction

The formulation of our objective function for motion estimation (Sect. 2.3) is based on a dynamic reconstruction algorithm $f(\boldsymbol{x}, \boldsymbol{r})$. The function f returns the reconstructed object value at a voxel $\boldsymbol{x}$ based on the motion model parameters $\boldsymbol{r}$. In principle, any dynamic reconstruction algorithm could be used. In this paper, an extension of the FDK reconstruction method for moving objets is utilized [3,7].

Formally, the dynamic FDK reconstruction $f : \mathbb{R}^3 \times \mathsf{P} \mapsto \mathbb{R}$ is given by

$$f(\boldsymbol{x}, \boldsymbol{r}) = \sum_i \underbrace{w(i, M(i, \boldsymbol{x}, \boldsymbol{r}))}_{=s_1} \cdot \underbrace{p(i, A(i, M(i, \boldsymbol{x}, \boldsymbol{r})))}_{=s_2}\,. \tag{2}$$

The function $w : \mathbb{N} \times \mathbb{R}^3 \mapsto \mathbb{R}$ is the distance weight of the FDK-formula. The pre-processed, filtered and redundancy weighted projection data is accessed by the function $p : \mathbb{N} \times \mathbb{R}^2 \mapsto \mathbb{R}$ where $p(i, \boldsymbol{m})$ returns the value of the i-th projection image at the pixel $\boldsymbol{m}$. The perspective projection is given by the function $A : \mathbb{N} \times \mathbb{R}^3 \mapsto \mathbb{R}^2$, where $A(i, \boldsymbol{x}) = \boldsymbol{m}$ maps a voxel $\boldsymbol{x}$ to a pixel location $\boldsymbol{m}$ in the i-th projection image. The shortcuts s_1, s_2 are introduced for convenience and are reused in Sect. 2.4.

2.3 Objective Function for Motion Estimation

The proposed approach for motion estimation is inspired by the works of Rohkohl et al. [7] and Hansis et al. [9]. Their results suggest that a sufficient initial reconstruction can be obtained by a FDK-like ECG-gated reconstruction which

can be improved by motion estimation and compensation [7,9]. Our method also requires such an inital reconstruction, which is post-processed such that only the sparse and contrasted vascular tree is visible. It is obtained by applying a transfer function in a volume renderer [7]. In the following the resulting reference image is denoted f_r. Regarding the quality of f_r it is especially important that larger structures are present. Due to the fact that the motion field is smooth in space and time, smaller structures are likely to move in accordance with the larger structures. Thus they will become visible in a motion corrected reconstruction.

Formally, motion estimation is formulated as a multi-dimensional optimization problem where the motion model parameters $\hat{\boldsymbol{r}} \in \mathsf{P}$ minimizing the objective function $\mathcal{O} : \mathsf{P} \mapsto \mathbb{R}$ need to be estimated, i.e.

$$\hat{\boldsymbol{r}} = \arg\min_{\boldsymbol{r}\in\mathsf{P}} \mathcal{O}(\boldsymbol{r}) = \arg\min_{\boldsymbol{r}\in\mathsf{P}} \left(\mathcal{L}(\boldsymbol{r}) + \mathcal{R}(\boldsymbol{r})\right), \tag{3}$$

with $\mathcal{R}$ being a regularizer and $\mathcal{L}$ being a dissimilarity measure. We decided for the dissimilarity function proposed by Rohkohl et al. [7]. It aims to maximize the joint intensity of the reference image f_r and another reconstruction – in our case the dynamic reconstruction $f(\boldsymbol{x}, \boldsymbol{r})$. It is given by

$$\mathcal{L}(\boldsymbol{r}) = -\sum_{\boldsymbol{x}} f_r(\boldsymbol{x}) \cdot f(\boldsymbol{x}, \boldsymbol{r}). \tag{4}$$

During the estimation of the motion model parameters, it needs to be ensured that only plausible motions are considered, i.e. no rapid motion or folding. Numerous regularizers have already been proposed for B-spline based motion models, e.g. [5,7] and hence will not be covered in this paper.

2.4 Optimization Strategy and Runtime Considerations

Many advanced optimization strategies require the gradient of (3), e.g. the L-BFGS-B algorithm [11]. The derivative of (3) with respect to the motion model parameters is given by

$$\frac{\partial \mathcal{O}(\boldsymbol{r})}{\partial \boldsymbol{r}} = \frac{\partial \mathcal{L}(\boldsymbol{r})}{\partial \boldsymbol{r}} + \frac{\partial \mathcal{R}(\boldsymbol{r})}{\partial \boldsymbol{r}} = -\sum_{\boldsymbol{x}} \left(f_r(\boldsymbol{x}) \frac{\partial f(\boldsymbol{x}, \boldsymbol{r})}{\partial \boldsymbol{r}} \right) + \frac{\partial \mathcal{R}(\boldsymbol{r})}{\partial \boldsymbol{r}}. \tag{5}$$

It contains the derivative $\frac{\partial f}{\partial \boldsymbol{r}}$ of the dynamic reconstruction (2) with respect to the motion model parameters. The derivative can be calculated analytically and is given by

$$\frac{\partial f(\boldsymbol{x}, \boldsymbol{r})}{\partial \boldsymbol{r}} = \sum_i \left(\frac{\partial w}{\partial M} s_2 + s_1 \frac{\partial p}{\partial A} \frac{\partial A}{\partial M} \right) \frac{\partial M(i, \boldsymbol{x}, \boldsymbol{r})}{\partial \boldsymbol{r}}. \tag{6}$$

The computation of the objective function and the corresponding derivative is fast as only non-zero voxel of the reference image f_r contribute to the objective function value. In most cases the number of non-zero voxels is $< 1\%$. This property can be exploited for obtaining a fast implementation. Further, the dynamic FDK reconstruction is highly parallelizable and can be efficiently implemented on modern hardware, e.g. the GPU [12].

3 Evaluation

3.1 Methods and Materials

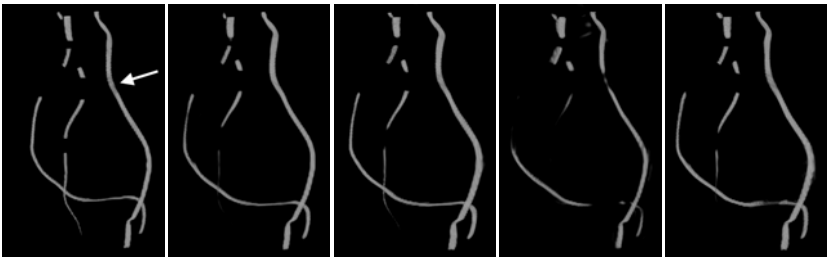
Evaluation Strategy. In order to assess the performance of the proposed algorithm, the reconstruction results of our method are compared to the approach of periodic motion [7]. In the periodic approach, ECG-gated reconstructions of several heart phases are registered with the reference image f_r by maximizing the joint intensity. For the presented results, ten disjoint ECG-gated subreconstructions have been performed using each 10% of all projection data. Both algorithms are provided with equivalent parameter settings. As optimizer, the L-BFGS-B algorithm [11] with a fixed number of 50 iterations was used. All datasets were acquired on an Artis zee C-arm system (Siemens AG, Healthcare Sector, Germany) covering an angular range of 200°.

Physical Heart Phantom. A physical heart phantom has been built from an elastic material filled with water. At a rate of 50 bpm the water is pumped in and out causing an elastic motion. In addition, the phantom is mounted on a compensator for a slow breathing motion of approx. 0.5 cycles per second. The artificial heart is surrounded by tubular structures filled with contrast agent mimicing the vascular system. An ECG-like signal is generated for the cardiac motion.

In total, three different datasets have been acquired. Each dataset consists of $N = 133$ projection images with an acquisition time of $T = 5$ seconds. The first dataset is the static phantom. It is used to determine the ground truth reconstruction. The second dataset is the moving phantom undergoing periodic cardiac motion. It is used to test the algorithms on cyclic motion patterns. The third dataset is the moving phantom undergoing a non-periodic combination of cardiac and respiratory motion. It is used to assess the algorithm performance on non-cyclic motion patterns.

For determing the quality of the reconstructions, a visual inspection of the volume renderings of all datasets using the same window settings, is performed. In addition, the vessel diameter, determined from MPR-views, is compared at a representative vessel position.

Clinical Cases: Arterial and Venous Vascular System. In order to demonstrate the clinical feasibility of the presented concepts, the algorithm is applied to the reconstruction of the venous and arterial vascular system. The first case is a coronary artery dataset with $N = 133$ and $T = 5$ seconds. The second case is a coronary sinus dataset with an $N = 397$ and $T = 8$ seconds. For the latter case, no ECG-signal is available due to a mislocation of the electrode. Periodic approaches cannot handle this situation. Our approach is feasible when using the non-gated reconstruction as reference image f_r.



(a) (b) (c) (d) (e)

Fig.	Description	Diameter	Abs. Error	Rel. Error
(a)	gold standard (no motion)	3.08 mm	-	-
(b)	periodic motion: classical method	3.24 mm	0.16 mm	5.19%
(c)	periodic motion: proposed method	3.24 mm	0.16 mm	5.19%
(d)	non-periodic motion: classical method	1.85 mm	1.23 mm	39.94%
(e)	non-periodic motion: proposed method	3.42 mm	0.34 mm	11.04%

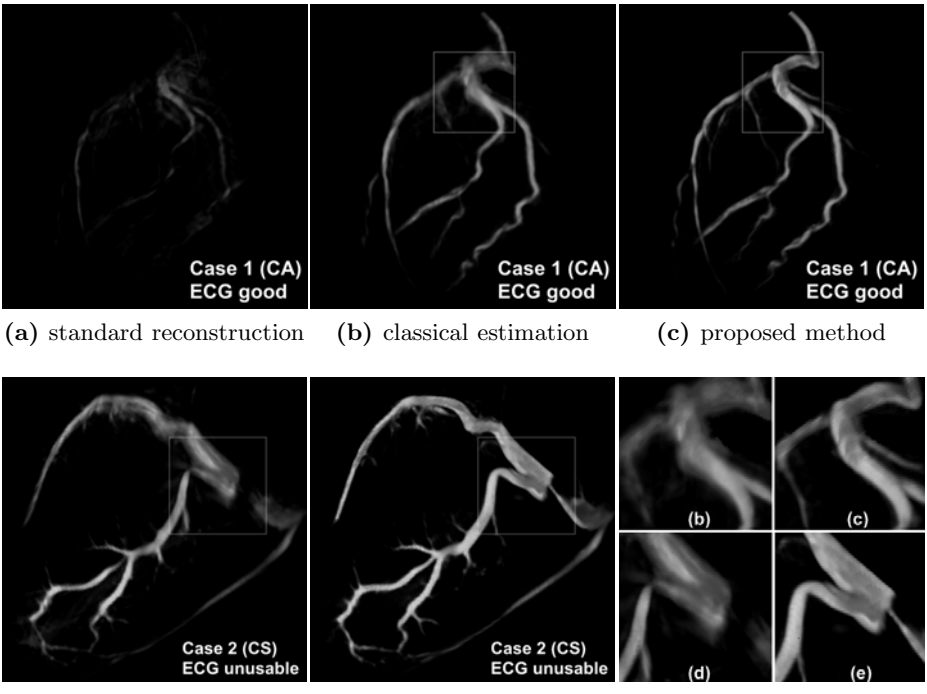
Fig. 1. Summary of the results of the phantom study. All volume renderings have common visualization settings. Details for the vessel diameter measurements and for the content of the images are listed in the table. The white arrow in the first image points to the measurement area of the vessel diameter.

3.2 Results and Discussion

Physical Heart Phantom. The volume rendering of the reconstruction results and the vessel diameter measurements are depicted in Fig. 1. For the case of periodic cardiac motion, both methods provide a similar reconstruction quality with a relative error of 5% in the vessel diameter. For the case of non-periodic motion, the proposed method clearly outperforms the periodic method. The relative error of the vessel diameter could be decreased by 29% from 40% (1.23 mm) to 11% (0.34 mm). This result is reasonable since our approach concerns the non-periodicity of the motion.

Clinical Cases: Arterial and Venous Vascular System. The volume rendering of the reconstruction results for the first case are shown in the upper row of Fig. 2. It can be seen that both motion compensated reconstruction methods are superior to the standard reconstruction without motion compensation. However, due to heart beat irregularities, our approach shows superior reconstruction results in terms of detail level and image artifacts.

The volume rendering of the reconstruction results for the second case are compared in the lower part of Fig. 2. It can be seen that the standard reconstruction shows blurring and artifical stenosises due to motion. The motion compensated reconstruction significantly reduces the motion artifacts and provides a higher level of detail. Due to the missing ECG signal no motion estimation and compensation could be performed with ECG-based perodic approaches.



(a) standard reconstruction **(b)** classical estimation **(c)** proposed method

(d) standard reconstruction **(e)** proposed method **(f)** details

Fig. 2. Results for two clinical cases. The upper row depicts the volume rendering of a coronary artery (CA) dataset with a good ECG-signal using three different reconstruction methods. The images (d) (e) in the second row depict a coronary sinus (CS) case without ECG-signal. In (f) a zoomed-in version of the areas indicated by a square is depicted.

4 Conclusion and Outlook

The presented approach improves the reconstruction quality of dynamic high contrast objects significantly. This is achieved by two main contributions. First, the peridocity assumption was abandoned and replaced by a time-continous 4-D motion model which is parameterized over the complete acquisition time. Second, a method which couples dynamic reconstruction and motion estimation was introduced. In our studies it proved to allow a robust estimation of the motion model parameters.

In a phantom and two clinical cases it could be shown that our method clearly improves the reconstruction quality in the presence of non-periodic motion and outperforms periodic approaches. In a phantom study the relative error of the vessel diameter could be decreased by 29% for non-cyclic motions.

The implications of our work for the development of future applications are diverse. Foremost, the reconstruction quality and stability in an interventional environment is increased. In long-term thinking, however, novel applications and

diagnostic tools may be developed. A full 4-D model of the cardiac vasculature motion could be computed for deriving functional information. E.g. cardiac dysfunctions could be detected by a 4-D animation over several heart beats. Further investigations will be subject of our upcoming research.

Disclaimer: The concepts and information presented in this paper are based on research and are not commercially available.

References

1. Strobel, N., Meissner, O., Boese, J., Brunner, T., Heigl, B., Hoheisel, M., Lauritsch, G., Nagel, M., Pfister, M., Rührnschopf, E.P., Scholz, B., Schreiber, B., Spahn, M., Zellerhoff, M., Klingenbeck-Regn, K.: 3D Imaging with Flat-Detector C-Arm Systems. In: Multislice CT, 3rd edn., pp. 33–51. Springer, Heidelberg (2009)
2. Desjardins, B., Kazerononi, E.: ECG-Gated Cardiac CT. American Journal of Roentgenology 182(4), 993–1010 (2004)
3. Schäfer, D., Borgert, J., Rasche, V., Grass, M.: Motion-compensated and gated cone beam filtered back-projection for 3-d rotational x-ray angiography. IEEE Transactions on Medical Imaging 25(7), 898–906 (2006)
4. Hansis, E., Schäfer, D., Grass, M., Dössel, O.: An iterative method for the reconstruction of the coronary arteries from rotational x-ray angiography. In: Hsieh, J., Flynn, M.J. (eds.) Proceedings of the SPIE Medical Imaging 2007: Physics of Medical Imaging, vol. 6510, p. 651026 (2007)
5. Blondel, C., Malandain, G., Vaillant, R., Ayache, N.: Reconstruction of coronary arteries from a single rotational x-ray projection sequence. IEEE Transactions on Medical Imaging 25(5), 653–663 (2006)
6. Prümmer, M., Wigstroem, L., Hornegger, J., Boese, J., Lauritsch, G., Strobel, N., Fahrig, R.: Cardiac C-arm CT: Efficient Motion Correction for 4D-FBP. In: Smith, G. (ed.) Nuclear Science Symposium, Medical Imaging, pp. 1–20 (2006)
7. Rohkohl, C., Lauritsch, G., Nöttling, A., Prümmer, M., Hornegger, J.: C-arm ct: Reconstruction of dynamic high contrast objects applied to the coronary sinus. In: IEEE NSS-MIC (Nuclear Science Symposium-Medical Imaging Conference), October 2008, p. M10-328 (2008)
8. Jandt, U., Schäfer, D., Grass, M., Rasche, V.: Automatic generation of time resolved 4d motion vector fields of coronary arteries. In: 9th Int. Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine, Lindau, Germany, pp. 249–252 (2007)
9. Hansis, E., Schäfer, D., Dössel, O., Grass, M.: Projection-based motion compensation for gated coronary artery reconstruction from rotational x-ray angiograms. Physics in Medicine and Biology 53(14), 3807–3820 (2008)
10. Unser, M.: Splines: A perfect fit for signal and image processing. IEEE Signal Processing Magazine 16(6), 22–38 (1999); IEEE Signal Processing Society's 2000 magazine award
11. Byrd, R.H., Lu, P., Nocedal, J., Zhu, C.: A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Comput. 16(5), 1190–1208 (1995)
12. Scherl, H., Keck, B., Kowarschik, M., Hornegger, J.: Fast GPU-Based CT Reconstruction using the Common Unified Device Architecture (CUDA). In: Frey, E.C. (ed.) Nuclear Science Symposium, Medical Imaging Conference 2007, pp. 4464–4466 (2007)

Estimating Continuous 4D Wall Motion of Cerebral Aneurysms from 3D Rotational Angiography

Chong Zhang[1,2], Mathieu De Craene[2,1], Maria-Cruz Villa-Uriol[1,2], Jose M. Pozo[1,2], Bart H. Bijnens[3,1,2], and Alejandro F. Frangi[1,2,3]

[1] Center for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB), Universitat Pompeu Fabra, Barcelona, Spain
[2] Networking Research Center on Bioengineering, Biomaterials and Nanomedicine (CIBER-BBN), Spain
[3] Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

Abstract. This paper presents a technique to recover dynamic 3D vascular morphology from a single 3D rotational X-ray angiography acquisition. The dynamic morphology corresponding to a canonical cardiac cycle is represented via a 4D B-spline based spatiotemporal deformation. Such deformation is estimated by simultaneously matching the forward projections of a sequence of the temporally deformed 3D reference volume to the entire 2D measured projection sequence. A joint use of two acceleration strategies is also proposed: semi-precomputation of forward projections and registration metric computation based on a narrow-band region-of-interest. Digital and physical phantoms of pulsating cerebral aneurysms have been used for evaluation. Accurate estimation has been obtained in recovering sub-voxel pulsation, even from images with substantial intensity inhomogeneity. Results also demonstrate that the acceleration strategies can reduce memory consumption and computational time without degrading the performance.

1 Introduction

One 3D rotational X-ray angiography (3DRA) acquisition provides a sequence of 2D rotational X-ray angiographic images and an isotropic high-resolution 3D volumetric image reconstructed from them. As the acquisition lasts for a few seconds, the vascular motion occurring during the acquisition is captured in the 2D projections. However, since only a single volume is reconstructed, it does not provide the motion in 3D. The objective of our work is to retrieve the dynamic 3D morphology of the imaged region from 3DRA by combining the volume, the projections, and an additionally synchronized physiological signal. The latter provides the temporal information of each projection in the cardiac cycles.

The application targeted in this paper is cerebral aneurysm wall motion estimation. Cerebral aneurysms are pathological dilations of brain arteries commonly located at the Circle of Willis. When they rupture, spontaneous subarachnoid hemorrhage usually follows, causing high morbidity and mortality rates [1].

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 140–147, 2009.

It has been observed that aneurysms pulsate over the cardiac cycle [2] and that areas of higher pulsation coincide with the rupture sites [3]. However, the recovery of such pulsation is challenging because it is expected to be in the order of the voxel size. Also, existing aneurysm wall motion studies suffer from either strong artifacts [3] or partial motion estimation due to the use of a single projection view [4,5]. In our previous work [6], the aneurysm morphology at a given time instant is estimated from its temporal vicinity by matching projections of a deformed 3D reference volume to a sparse set of 2D projections in a weighted scheme. Similar to traditional gated reconstruction techniques, this approach approximates the spatiotemporal motion independently from one discrete time point to another, but fails to fully address the nature of motion: temporal consistency, that is, continuity over time.

In this paper, temporal consistency is achieved by formulating the inputs from 3DRA acquisitions into a 4D to multiple 2D image registration framework, which can be also found in [7]. However, we propose to build a canonical cardiac cycle and then match the forward projections of the temporally deformed 3D reference volume to the 2D measured projection sequence. As a result, the dynamic morphology of cerebral aneurysms could be described continuous and smooth both spatially and temporally. However, the cost of memory storage and computational time for processing a high-resolution 4D image and two large 2D image sequences at the same time is very high. We therefore further introduce a joint use of two acceleration strategies: semi-precomputation of forward projections and registration metric computation based on a narrow-band region-of-interest (ROI).

2 Algorithm Framework

Building a Canonical Cardiac Cycle. During the rotation of the C-arm gantry, the total angular coverage of the measured projections during each cardiac cycle is 40-50°. This small range of the projection views may be less informative about the motion along certain directions. This could be compensated by introducing an additive constraint to the optimization process as in [7]. We overcome this limitation by reordering all the projections into one canonical cardiac cycle. Thus the projection viewing angle range in any temporal vicinity is enriched, which subsequently ensures consistent 3D morphology estimation from all directions. An illustration of this process is shown in Fig. 1(a). We first normalize the period of each cardiac cycle to 1, according to the synchronized physiological signal, e.g. electrocardiogram or pressure waveform. Hence any projection is acquired within the $[0, 1)$ interval. All the N projections are then sorted by this normalized time to build one canonical cardiac cycle as $\mathcal{I} = \{I_{t_k}(\mathbf{x}) \mid k = 1 \dots N\ ,\ \mathbf{x} \in \mathcal{S}_k \subset \mathbb{R}^3\}$, where $I_{t_k}(\mathbf{x})$ represents the measured projection with the rotating field of view at projection plane $\mathcal{S}_k$ and at time t_k, $0 \leq t_k \leq t_{k+1} < 1$.

4D to Multiple 2D Image Registration. As shown in Fig. 1(b), motion throughout the cardiac cycle is estimated by a 4D deformation $\mathcal{T}$ with parameters $\boldsymbol{\omega}$. And the 3D instantaneous morphology at time t is given by:

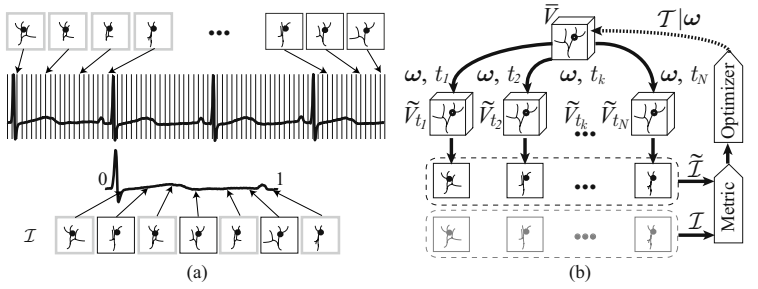


Fig. 1. (a) Building a canonical cardiac cycle. (b) 4D to multiple 2D image registration.

$$\widetilde{V}_t(\mathbf{p}) = \overline{V}(\mathcal{T}(\boldsymbol{\omega}, \mathbf{p}, t)), \tag{1}$$

where $\mathbf{p}$ is a point on $\widetilde{V}_t$, and $\overline{V}$ stands for the reference volume. The deformation function $\mathcal{T}$ is chosen to be a B-spline based transformation [8]:

$$\mathcal{T}(\boldsymbol{\omega}, \mathbf{p}, t) = \mathbf{p} + \sum_{\tau} \sum_{\mathbf{c}} \beta\left(\frac{t - t^{\tau}}{\Delta^{\tau}}\right) B\left(\frac{\mathbf{p} - \mathbf{p}^{\mathbf{c}}}{\boldsymbol{\Delta}^{\mathbf{c}}}\right) \boldsymbol{\omega}^{\tau,\mathbf{c}}, \tag{2}$$

where $B(\cdot)$ is the 3D tensor product of 1D cubic B-spline functions $\beta(\cdot)$, defined on a sparse grid of 4D control points $(\mathbf{p}^{\mathbf{c}}, t^{\tau})$, being $\mathbf{c}$ the spatial index and τ the temporal index, and $(\boldsymbol{\Delta}^{\mathbf{c}}, \Delta^{\tau})$ the width of the functions in each dimension.

For each I_{t_k} a corresponding forward projection $\tilde{I}_{t_k}$, also called digitally reconstructed radiograph (DRR), is generated to approximate the X-ray angiography. The pixel value $\tilde{I}_{t_k}(\mathbf{x})$ is calculated by integrating the voxel values of $\widetilde{V}_{t_k}$ along the ray from the X-ray source $\mathbf{s}$ to $\mathbf{x}$ in the projection plane (see Fig. 2(a)):

$$\tilde{I}_{t_k}(\mathbf{x}) = \int_{\mathcal{L}_{\mathbf{s},\mathbf{x}}} \widetilde{V}_{t_k}(\boldsymbol{\gamma}) d\boldsymbol{\gamma} \approx \sum_{\mathbf{p}_\ell \in \mathcal{L}_{\mathbf{s},\mathbf{x}}} \alpha_\ell \, \overline{V}(\mathcal{T}(\boldsymbol{\omega}, \mathbf{p}_\ell, t_k)), \tag{3}$$

where p_ℓ denotes the ℓth sampled point along $\mathcal{L}_{\mathbf{s},\mathbf{x}}$, and α_ℓ the sampling weight, which is the distance between two consecutive sampled points $\mathbf{p}_\ell$ and $\mathbf{p}_{\ell+1}$. Similarly, we denote $\tilde{\mathcal{I}}$ as the DRRs sequence, which is iteratively modified to match the measured projection sequence $\mathcal{I}$ for an optimal estimation with $\hat{\boldsymbol{\omega}}$:

$$\hat{\boldsymbol{\omega}} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \left\{ M(\boldsymbol{\omega}, \mathcal{I}, \tilde{\mathcal{I}}) \right\}, \tag{4}$$

where M is the similarity metric between two mapping regions. The L-BFGS-B algorithm [9] was used as the optimizer, and mutual information [10] as the metric function. Instead of computing the histograms individually for each projection pair, all sampled points from the entire sequence are considered as one region, thus forming a single histogram. That is, M describes the similarity between two sequences. After obtaining $\hat{\boldsymbol{\omega}}$, we could substitute it into Eq. 1 to extract instantaneous motion at any given time instant.

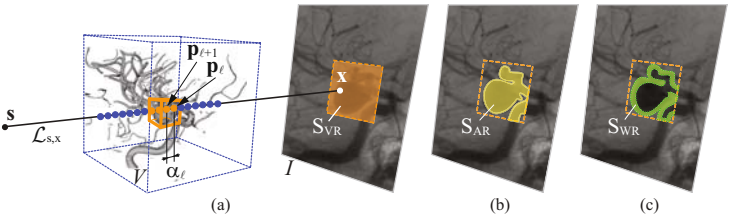


Fig. 2. (a) An illustration of DRRs computation. The sampling region S_{VR} is the projected volume-of-interest. We introduce two object-adaptive sampling regions: (b) the aneurysm region S_{AR} and (c) the aneurysm wall region S_{WR}.

3 Acceleration Strategies

Dealing simultaneously with a high-resolution 4D image and two large 2D image sequences, however, suffers from excessive memory requirements and computational load. For the method to be practically applicable, reducing both of them without degrading the performance is highly desirable. The following two joint strategies serve this purpose.

Semi-precomputed DRRs. A typical example of 3DRA patient data is shown in Fig. 2(a), where the structure of interest (i.e. aneurysm) is in the order of millimeters, occupying a very small part in the image. Thus, only this volume of interest (VOI) is deformed and the rest of the volume remains unmodified. On the other hand, in order to simulate X-ray projections, voxels of the entire volume must be integrated at each iteration to update forward projections. In order to avoid repeated computation, we precompute the voxel intensity integrals outside the VOI at each projection view. For instance in Fig. 2(a), the changing intensity sum of the sampled points on the ray $\mathcal{L}_{\mathbf{s},\mathbf{x}}$ in the VOI is added at each iteration to the precomputed sum of the remaining ones for the pixel value at $\mathbf{x}$.

Narrow-Band Region-of-interest. Typically the projected VOI, denoted as S_{VR} (Fig. 2(a)), is used as the sampling region to compute the metric. In this paper, we introduce two object-adaptive sampling regions: the projected aneurysm (S_{AR}) and the projected aneurysm wall (S_{WR}), shown in Fig. 2(b) and (c), respectively. Hence the image sequences in Eq. 4 are replaced by the sequences of these sampling regions. According to a typical histogram of a 3DRA cerebrovascular image, vessels belong to high intensity regions, among which the aneurysm region has even higher intensities [11]. This property allows to roughly discriminate different regions based on intensity. S_{AR} is obtained by first selecting a typical voxel value of the aneurysm dome, and then projecting only the points with voxel values larger than it. Similarly, S_{WR} is obtained by taking the non-shared region of the two projected regions from typical voxel values of the background and the dome. Currently, we assume for each individual image the typical intensity values of the background and the aneurysm dome by sampling two small patches from these regions, respectively. Additionally, using S_{AR} instead of S_{VR} is to avoid possible influences of background structures on the estimation, and

using S_{WR} is to further avoid potential registration failures caused by intensity inhomogeneity at the projected aneurysm dome.

4 Experiments and Results

Phantom and Image Acquisition. A typical 3DRA flat panel system like Allura Xper FD20 (Philips Healthcare, Best, The Netherlands) acquires 120 X-ray angiography images (resolution 1024^2 pixels with $0.15^2 mm^2$/pixel), during a four-second C-arm rotation of over 200 degrees. The images are then used to reconstruct a 3D volume of 256^3 voxels at resolutions of $0.3^3 mm^3$. In our experiments, 12 cases of digital phantom data were simulated to be consistent with these imaging settings using computer synthesized irregular geometries of dome diameter=8, 10, and 12mm and artery diameter=4mm. Maximum pulsation amplitudes were 1%-4% of the dome diameter (0.08-0.48mm). In order to simulate realistic 3DRA data, we fused the phantom data into a clinical dataset (Fig. 3). Details of this simulation are described in [6]. Physical phantom data was obtained with a silicone side-wall aneurysm phantom (Elastrat, Geneva, Switzerland), representing realistic dimensions of vascular structures in human bodies: dome diameter=10mm and artery diameter=4mm. It was filled with contrast agent (Iomeron 400, Bracco Imaging SpA, Milan, Italy) and connected to a customized pulsatile pump. Images were acquired from an aforementioned Allura Xper FD20 imaging suite.

Results. Experiments were executed on an Intel ® Core™2 Quad CPU Q6600 2.40GHz with 4GB of memory and no parallelization was used. All the experiments employed the semi-precomputed DRRs strategy, which essentially produced equal results but with a 3-4x speedup. This ratio is the proportion between the ray path length of the full volume and that of the VOI. The B-spline control point grid spacing was about 1.5mm for the spatial dimensions, and 12.5% of the canonical cardiac cycle for the temporal dimension.

We first carried out digital phantom experiments for quantitative validation. A set of deformed volume images at discrete time points were extracted according to the estimated transformation. Then, a relative estimation error was measured at each time point t as a percentage of the pulsation range: $\varepsilon(t) = \big(m_r(t) - m_g(t)\big)/\hat{m}_g \times 100\%$, where $m_g(t)$ is the ground-truth pulsation measurement (e.g., volume) at t, $m_r(t)$ the corresponding estimated measurement, and $\hat{m}_g$ the variation range of $m_g(t)$ over the canonical cycle. For each of the 12 cases in Fig. 3, a boxplot is shown, indicating the relative error in volume changes, ε_V, at 16 equally distributed time points. Aneurysm volume was measured using a similar strategy as in [12]. Except for two tests with maximum pulsation less than 0.1mm, estimation errors were below 10%.

The following experiment aimed at analyzing the sensitivity to spurious projection intensity variations. Although it is mainly caused by blood turbulence during contrast agent filling, the instantaneous local inhomogeneity is caused by multiple factors. We have synthesized such pattern based on a clinical dataset that presented substantial nonuniform contrast filling. For each of the digital

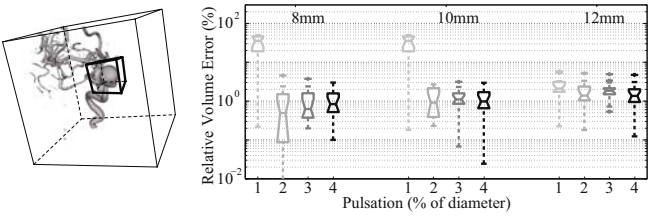


Fig. 3. Boxplots of ε_V at 16 equally distributed time points for 12 digital phantom cases, in which a digital phantom is fused into a clinical dataset

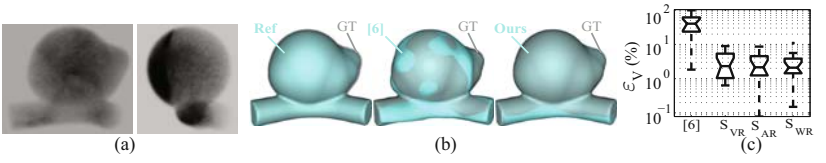


Fig. 4. (a) Projections with contrast inhomogeneity synthesized based on clinical data. (b) The ground-truth shape (GT) at one time instant overlaid with (from *left* to *right*): the reference shape (Ref), estimation using the method in [6], and our estimation. (c) Comparison of ε_V between [6] and our technique with three sampling regions.

phantom projections, an image patch was taken from the aneurysm dome of one different clinical projection and texture-mapped to the projected phantom dome, see Fig. 4(a). Then, we performed motion estimation using our technique with regions S_{VR}, S_{AR}, and S_{WR}, and the method in our previous work [6]. As shown in Fig. 4(b) and (c), a relative volume error ε_V below 10% was achieved using our technique, whereas the method in [6] failed to properly retrieve the motion. This suggests that our 4D smooth estimation is more robust to large image intensity inhomogeneity both temporally and spatially. Additionally, a slightly higher accuracy was obtained using the narrow-band ROI as most of the inhomogeneous regions did not contribute to the metric computation.

In terms of computational efficiency, we investigated the effects of using different sampling regions and angular resolutions along the C-arm gantry trajectory. An example for a phantom having 10-mm diameter and 3% maximum pulsation is given in Fig. 5, where ε_V is plotted together with the computational time for the same number of iterations. The angular resolution of the measured projection sequence was downsampled by a factor of 1-4. Results show similar accuracy (<5%) achieved from the three sampling regions combined with the angular resolution downsampling factor up to 3. However, the computational time can be proportionally reduced due to both of them. Compared to the computational time of 10-15 minutes for an estimation at a single time instant in our previous work [6], we could obtain the motion over the whole cycle in 15-20 minutes using S_{WR} with the angular resolution downsampling factor of 3. As a result, the use of the narrow-band ROI combined with a subsampled projection set could potentially save both memory and computational time without degrading the performance.

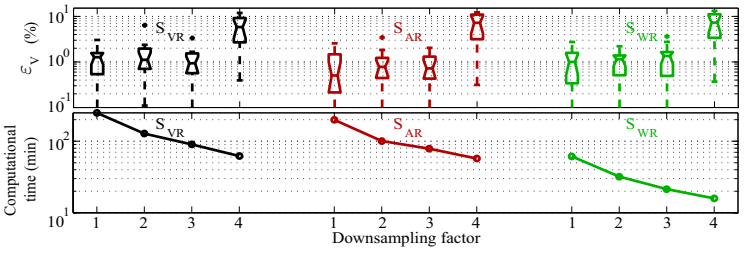


Fig. 5. Comparison in estimation error ε_V and computational time using different sampling regions and angular resolution downsampling factors

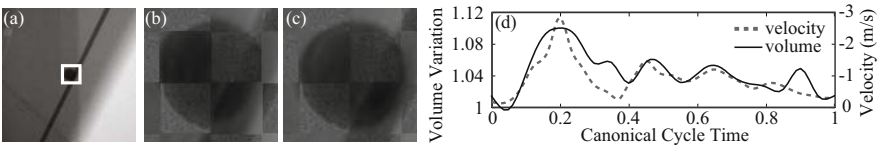


Fig. 6. (a) An X-ray projection image. (b,c) Checkerboard images of the measured and the forward projection from the reference volume and our estimation, respectively. (d) The volume variations along with the Doppler ultrasound velocity waveform.

We further tested the proposed technique on 3DRA data of the pulsating physical phantom acquired in a clinical environment. Due to the unknown ground-truth, results are qualitatively evaluated. As shown in Fig. 6, our estimation was able to correct the misalignment between the measured and the forward projections. The volume variation waveform was also compared to a Doppler ultrasound waveform measured at the parent vessel inlet. A similar pattern is observed between them, which is in agreement with the findings in [13].

5 Discussion and Conclusions

We have demonstrated on digital and physical phantoms the performance of the proposed technique in estimating continuous 4D cerebral aneurysm wall motion from 3DRA. Accurate estimation has been obtained in recovering sub-voxel pulsation, even from images with substantial intensity inhomogeneity. The joint use of the acceleration strategies has shown a reduction of memory cost as well as computational time (3-4x speedup from semi-precomputation and 10x from narrow-band ROI), while preserving the accuracy of the motion estimation. These results suggest its feasibility towards estimating 4D aneurysmal wall motion from patient data. Finally, the fact that motion is estimated from a single 3DRA acquisition dataset implies no additional radiation risk for the patient, which facilitates its clinical take-up.

Acknowledgments. The authors thank Dr. D. Rüfenacht and L. Augsburger from Geneva University Hospitals, for providing the aneurysm silicone phantom

and the pulsatile pump, and Dr. J. Blasco, Dr. J. Macho and Dr. E. Vivas for the 3DRA acquisitions of the physical phantom. This work was partially supported by Philips Healthcare, Best, The Netherlands, the CENIT-CDTEAM grant funded by the Spanish Ministry Of Science and Innovation (MICINN-CDTI), and partially generated in the framework of the @neurIST Integrated Project, which is co-financed by the European Commission (IST-027703).

References

1. Brisman, J.L., Song, J.K., Newell, D.W.: Medical progress: cerebral aneurysms. N. Engl. J. Med. 355(9), 928–939 (2006)
2. Meyer, F.B., Huston III, J., Riederer, S.S.: Pulsatile increases in aneurysm size determined by cine phase-contrast MR angiography. J. Neurosurg. 78(6), 879–883 (1993)
3. Ishida, F., Ogawa, H., Simizu, T., Kojima, T., Taki, W.: Visualizing the dynamics of cerebral aneurysms with four-dimensional computed tomographic angiography. Neurosurgery 57(3), 460–471 (2005)
4. Dempere-Marco, L., Oubel, E., Castro, M., Putman, C., Frangi, A.F., Cebral, J.R.: CFD analysis incorporating the influence of wall motion: Application to intracranial aneurysms. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 438–445. Springer, Heidelberg (2006)
5. Oubel, E., De Craene, M., Putman, C., Cebral, J., Frangi, A.F.: Analysis of intracranial aneurysm wall motion and its effects on hemodynamic patterns. In: Proc. SPIE Med. Imag.: Physiology, Function, and Structure from Medical Images, p. 65112A (2007)
6. Zhang, C., Villa-Uriol, M.-C., De Craene, M., Pozo, J.M., Frangi, A.F.: Morphodynamic analysis of cerebral aneurysm pulsation from time-resolved rotational angiography. IEEE Trans. Med. Imaging 28(7), 1105–1116 (2009)
7. Zeng, R., Fessler, J.A., Balter, J.M.: Estimating 3-D respiratory motion from orbiting views by tomographic image registration. IEEE Trans. Med. Imaging. 26(2), 153–163 (2007)
8. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Trans. Med. Imaging 18(8), 712–721 (1999)
9. Zhu, C., Byrd, R.H., Nocedal, J.: L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization. ACM Trans. Math. Software 23(4), 550–560 (1997)
10. Mattes, D., Haynor, D.R., Vesselle, H., Lewellen, T.K., Eubank, W.: PET-CT image registration in the chest using free-form deformation. IEEE Trans. Med. Imaging. 22(1), 120–128 (2003)
11. Bogunović, H., Radaelli, A.G., De Craene, M., Delgado, D., Frangi, A.F.: Image intensity standardization in 3D rotational angiography and its application to vascular segmentation. In: Proc. SPIE Med. Imag.: Image Processing, p. 691419 (2008)
12. Holden, M., Schnabel, J.A., Hill, D.L.G.: Quantification of small cerebral ventricular volume changes in treated growth hormone patients using nonrigid registration. IEEE Trans. Med. Imaging. 21(10), 1292–1301 (2002)
13. Löw, M., Perktold, K., Raunig, R.: Hemodynamics in rigid and distensible saccular aneurysms: a numerical study of pulsatile flow characteristics. Biorheology 30, 287–298 (1993)

NIBART: A New Interval Based Algebraic Reconstruction Technique for Error Quantification of Emission Tomography Images

Olivier Strauss[1], Abdelkabir Lahrech[1], Agnès Rico[2], Denis Mariano-Goulart[3], and Benoît Telle[1]

[1] LIRMM UMR CNRS 5506, Université Montpellier 2, 161, rue Ada 34392 Montpellier cedex 5, France
strauss@lirmm.fr
[2] LIRIS UMR CNRS 5205, Université Claude Bernard Lyon 1, 43 avenue du 11 novembre 1918, 69622 Villeurbanne, France
[3] CHU Lapeyronie, Service Central de Médecine Nucléaire, 371, avenue du doyen Gaston Giraud 34295 Montpellier cedex 5, France

Abstract. This article presents a new algebraic method for reconstructing emission tomography images. This approach is mostly an interval extension of the conventional SIRT algorithm. One of the main characteristic of our approach is that the reconstructed activity associated with each pixel of the reconstructed image is an interval whose length can be considered as an estimate of the impact of the random variation of the measured activity on the reconstructed image. This work aims at investigating a new methodological concept for a reliable and robust quantification of reconstructed activities in scintigraphic images.

1 Introduction

In the field of nuclear medicine, the need for a reliable comparison of reconstructed activities in two regions of interest often occurs when the question is to decide whether the increase in the metabolism of a tissue is high enough, compared with normal tissue in the neighborhood, to diagnose or to exclude a particular disease. Since any reconstruction method tends to highly correlate the acquisition error, a robust comparison of reconstructed activities in two (or more) regions of interest leads to the need for a general methodology for reliably predicting the noise and resolution properties in reconstructed images.

Several solutions have been proposed in the relevant literature for quantifying the noise in the reconstructed slices. They generally assume that the reconstructed image results from some optimization of a functional that takes into account both a precise geometric description of the tomographic problem, a model of the statistical properties of the measured projections and some a-priori knowledge on the resolution of image to be reconstructed [1,2,3,4]. All these works are based on an empirical or a numerical inversion of the Jacobian of the functional, that is minimized by the reconstruction problem, to analytically

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 148–155, 2009.

deduce the statistical properties of the reconstructed image from a presupposed Poisson noise in the projections. Such an inversion is hard to compute with accuracy for the ill-posed problem of tomography, especially when large image sizes are used. This makes necessary to simplify the projection process, thus leading to results whose accuracy and reliability is badly known and sub-optimal. Even when strong simplifying hypotheses are put forward, the proposed methods are generally too computationally expensive[1,2,3]. In [4], Budinger et al. propose an empirical method for predicting the noise but this method concerns only reconstruction algorithms that involve a single back-projection procedure (so mostly filtered back projection).

In this article, we develop an efficient and faster approach for estimating the image reconstruction error. This approach is based on considering a new projection operator whose output is interval valued. As shown in [5], the imprecision of this output interval is highly correlated with the statistical impact of the random variations of the measures. In the present work, we keep these quantification properties by extending a classical algebraic reconstruction technique. The error quantification ability of the proposed technique is highlighted through a simulated experiment.

This paper will be structured as follows: Section 2 briefly describes the Simultaneous Iterative Reconstruction Technique (SIRT). In Section 3, we recall some interval arithmetic operations. Section 4 presents the new interval valued projection operator we use. Section 5 describes the NIBART approach. Section 6 presents an illustrative experiment highlighting the ability of the NIBART approach to provide a noise quantification.

2 Algebraic Reconstruction Techniques

The reconstruction problem in emission tomography consists in determining the activity distribution in each pixel of the image to be reconstructed by using noisy discrete observations of its projection. Let $\mathbf{I} = (I_1, \ldots, I_N)$ be the activity vector of the N pixels to be reconstructed and $\mathbf{S} = (S_1, \ldots, S_K)$ denote the measured activity associated to each of the K detector positions. If no noise is assumed in the acquisition process, then the relationship between $\mathbf{S}$ and $\mathbf{I}$, known as the discrete Radon transform, is defined by:

$$S_k = \sum_{n=1}^{N} \rho_{k,n} I_n = \mathcal{P}(\mathbf{I}), \tag{1}$$

where $\mathcal{P}$ is a linear projection operator based on a $K \times N$ Radon matrix. Each $\{\rho_{k,n}\}$ element of the Radon matrix can be considered as the probability of the activity in the n^{th} pixel to be detected by the k^{th} detector. Reconstructing the activity distribution $\mathbf{I}$ from the measurement distribution $\mathbf{S}$ consists in inverting the hight dimensional ill-conditioned linear equation Eq.(1). Many algebraic methods have been proposed in the relevant literature that consists in iteratively modifying a guessed activity density to bring each reconstructed projection in

agreement with the measured projections. The iterative modification can be additive (see e.g. [6]) or multiplicative (see e.g. [7,8]). In this paper, we consider the SIRT method which achieves a global additive modification of the activity because its interval extension rather straightforward. Conventionally, the SIRT algorithm can be defined by the following equation [9]:

$$\forall n \in \{1,\ldots,N\}, \quad I_n^i = I_n^{i-1} + \lambda \sum_{k=1}^{K} \frac{\rho_{k,n}}{\sum_{j=1}^{K} \rho_{j,n}} \left(S_k - \sum_{m=1}^{N} \rho_{k,m} I_m^{i-1} \right), \tag{2}$$

where i is the iteration number and $\lambda \in [0,1]$ is a relaxation parameter.

For sake of simplicity, the SIRT algorithm (see Eq.(2)) can be reformulated as :

$$\mathbf{I}^i = \mathbf{I}^{i-1} + \lambda \mathcal{B}^* \left(\mathbf{S} - \mathcal{P}(\mathbf{I}^{i-1}) \right), \tag{3}$$

where $\mathcal{B}^*$ is the linear normalized back-projection operator based on the transposed Radon matrix. Using a SIRT-like algorithm to solve Eq.(1) can need a very large number of iterations to complete the reconstruction. Though, when used with noisy emission tomography measurements, the iterative reconstruction process leads to images with spurious features that rises owing to the ill-conditionnedness of the problem. Moreover, the level of these spurious features increases with the number of iterations. This last drawback can be a severe limit for using such a method since the few techniques that are available to provides a reliable termination criterion of the algorithm are rather intricate and generally not used in clinical routine [10].

3 Interval Arithmetic

In this section, we briefly present the interval arithmetic operations that will be used to extend the conventional SIRT to interval valued data. In the following, $[a] = [\underline{a}, \overline{a}]$ will denote a real interval whose lower (rsp. upper) value is $\underline{a}$ (rsp. $\overline{a}$). Let $[a]$ and $[b]$ be two real intervals, the extension of addition to intervals can be achieved in two ways [11]:

- the Minkowski addition:

$$[a] \oplus [b] = [\underline{a}, \overline{a}] \oplus [\underline{b}, \overline{b}] = [\underline{a} + \underline{b}, \overline{a} + \overline{b}], \tag{4}$$

- the dual Minkowski addition:

$$[a] \boxplus [b] = [\underline{a}, \overline{a}] \boxplus [\underline{b}, \overline{b}] = [min(\underline{a} + \overline{b}, \overline{a} + \underline{b}), max(\underline{a} + \overline{b}, \overline{a} + \underline{b})]. \tag{5}$$

The extension of subtraction is simply obtained by: $[a] \ominus [b] = [a] \oplus (-[b])$ and $[a] \boxminus [b] = [a] \boxplus (-[b])$, $(-[b])$ being equal to $[-\overline{b}, -\underline{b}]$.

Note that only one of the two equations $[x] \ominus [b] = [a]$ and $[x] \ominus [a] = [b]$ has a solution and that this solution is $[x] = [a] \boxplus [b]$. Moreover, the following inclusion is always true: $([a] \boxplus [b]) \subseteq ([a] \oplus [b])$.

The extension of these operations to interval valued vectors is straightforward. Let $[\mathbf{A}] = [\underline{\mathbf{A}}, \overline{\mathbf{A}}] = ([a_1], \ldots, [a_N])$ and $[\mathbf{B}] = [\underline{\mathbf{B}}, \overline{\mathbf{B}}] = ([b_1], \ldots, [b_N])$ be two N-dimension interval valued vectors, the arithmetic operation $\circ$ of $[\mathbf{A}]$ and $[\mathbf{B}]$ is defined by:

$$[\mathbf{A}] \circ [\mathbf{B}] = ([a_1] \circ [b_1], \ldots, [a_N] \circ [b_N]), \tag{6}$$

with $\circ \in \{\oplus, \ominus, \boxplus, \boxminus\}$.

4 A New Projection Operator

In a recent paper [5] a new way for modeling the projection operators commonly used in classical reconstruction algorithms has been proposed. This approach is based on an attempt to overcome the classical problem of discretizing the continuous Radon transform by directly using a proper discrete transform called the Hough transform. Applying the Hough transform formalism to emission tomography induces a non-additive confidence measure, instead of the conventional probabilistic confidence measure, to compute the relation between the activity of a region of the image and the observed values measured by the tomographic device. As an important consequence of this replacement, the output of the projection operator is interval-valued. Moreover, as shown in [5], the length of the interval valued activity is highly correlated with the level of its random variation. In the rest of the paper, the interval valued projection operator will be denoted $\overline{\underline{\mathcal{P}}}$:

$$[\mathbf{S}] = \overline{\underline{\mathcal{P}}}(\mathbf{I}), \tag{7}$$

$[\mathbf{S}]$ being the interval valued projection vector and $\mathbf{I}$ the activity vector. The interval valued projection $[\mathbf{S}]$ can be thought as the convex set of all the values that would be obtained by using a coherent family of linear projector $\mathcal{P}$, coherent meaning that they are based on the same geometrical model. We will denote $\mathcal{F}(\overline{\underline{\mathcal{P}}})$ this family: $\mathcal{F}(\overline{\underline{\mathcal{P}}}) = \{\mathcal{P} / \forall \mathbf{I} \in \mathbb{R}^N, \mathcal{P}(\mathbf{I}) \in \overline{\underline{\mathcal{P}}}(\mathbf{I})\}$. Since the operator $\overline{\underline{\mathcal{P}}}$ preserves the order, this interval valued projector can be easily extended to an interval valued input: $[\mathbf{S}] = \overline{\underline{\mathcal{P}}}([\mathbf{I}]) = [\underline{\mathbf{S}}, \overline{\mathbf{S}}]$, with $[\mathbf{I}] = [\underline{\mathbf{I}}, \overline{\mathbf{I}}]$, $\underline{\mathbf{S}}$ being the lower value of $\overline{\underline{\mathcal{P}}}(\underline{\mathbf{I}})$, $\overline{\mathbf{S}}$ being the upper value of $\overline{\underline{\mathcal{P}}}(\overline{\mathbf{I}})$. When considering the same geometrical modeling, an associated interval valued back-projection operator $\overline{\underline{\mathcal{B}}}$ can be also easily derived which represents the set of all the back-projection operators associated to each of the projectors of the family $\mathcal{F}(\overline{\underline{\mathcal{P}}})$.

The most relevant explanation for the noise quantifying ability of this interval-based projection operator is the following. $\mathcal{F}(\overline{\underline{\mathcal{P}}})$ is a coherent subset of linear projector that should provide the same output when there is no random variation in the data to be projected. When the data are noisy, then the dispersion of the different output values given by using each $\mathcal{P} \in \mathcal{F}(\overline{\underline{\mathcal{P}}})$ is a marker of the impact of the noise level on the projection to be performed.

5 A New Interval Based Algebraic Reconstruction Technique (NIBART)

The challenge of NIBART is to provide an interval valued reconstructed activity that keeps the noise quantification ability of the interval valued projection and back projection operators presented in Section 4. Roughly speaking, NIBART can be thought as a simple interval-based extension of the SIRT method displayed by Eq.(3). In fact, according to Eq.(3), at the i^{th} iteration, the right additive correction to apply to the guessed image $\mathbf{I}^{i-1}$, for making an image $\mathbf{I}^i$ whose projection is closer from the measured projection $\mathbf{S}$, is likely to be $\mathbf{D}^i = \lambda \mathcal{B}^* \left(\mathbf{S} - \mathcal{P}(\mathbf{I}^{i-1})\right)$. In short, $\mathbf{D}^i$ is the expected value of $(\mathbf{I}^{i-1} - \mathbf{I}^i)$.

Now, when considering the interval-valued projector and back-projector and an interval valued previous guess $[\mathbf{I}^{i-1}]$, the interval valued vector $[\mathbf{D}^i]$ defined by $[\mathbf{D}^i] = \lambda \underline{\mathcal{B}^*} \left([\mathbf{S}] \ominus \overline{\mathcal{P}}([\mathbf{I}^{i-1}])\right)$ is the convex subset of all the additive corrections that can be applied to one $\mathbf{I} \in [\mathbf{I}^{i-1}]$ when considering one $\mathcal{P} \in \mathcal{F}(\overline{\mathcal{P}})$ and one $\mathcal{B} \in \mathcal{F}(\overline{\mathcal{B}})$. The NIBART challenge is now turned into finding the most specific interval $[\mathbf{I}^i]$ that is in agreement with the equation $[\mathbf{D}^i] = [\mathbf{I}^i] \ominus [\mathbf{I}^{i-1}]$. As mentioned in Section 3, this interval is given by:

$$[\mathbf{I}^i] = [\mathbf{I}^{i-1}] \boxplus [\mathbf{D}^i] = [\mathbf{I}^{i-1}] \boxplus \lambda \underline{\mathcal{B}^*} \left(\mathbf{S} \ominus \overline{\mathcal{P}}([\mathbf{I}^{i-1}])\right). \tag{8}$$

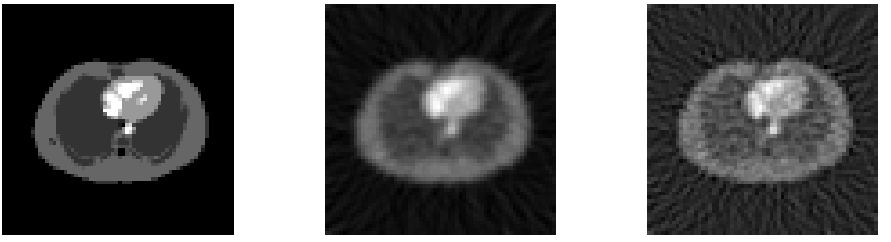
Note that the precise vector $\mathbf{S}$ in expression (8) can be replaced by an interval valued vector $[\mathbf{S}]$ to account for a known error in the measurement (e.g. $[\mathbf{S}]$ is the 99% confidence interval of the measurement vector $\mathbf{S}$). Within the NIBART method, a certain convergence criterion can be detected since empirical results show that, after a few iterations, the measured interval $[\mathbf{S}]$ intersects the projected interval $\overline{\mathcal{P}}([\mathbf{I}^i])$. This intersection expresses the fact that there is at least one value $\mathbf{S} \in [\mathbf{S}]$, one value $\mathbf{I} \in [\mathbf{I}^i]$ and one projector $\mathcal{P} \in \mathcal{F}(\overline{\mathcal{P}})$ such that $\mathbf{S} = \mathcal{P}(\mathbf{I})$. We call it the weak adequacy.

What has been settled in the many experiments we carried out, and that will be illustrated in the next section, is that, whatever the convergence conditions, the reconstruction error due to statistical noise in the measurements is highly correlated with the length of the reconstructed interval valued activity.

6 Experiments

This section aims at illustrating the ability of the NIBART method to provide a reliable quantification of the reconstruction error induced by the statistical noise in the measurements, whatever the convergence conditions. To achieve this experiment, we have simulated 100 noisy projections of a 64×64 Zubal phantom[1] displayed in Fig.1(a). These noisy projections have been obtained by corrupting a noise free projection of the phantom (total count per projection $\approx$ 10 kcounts) with a Poisson process. The 100 simulated 64×64 noisy sinograms will be considered as 100 realizations of the same projection. For each

[1] Available at http://noodle.med.yale.edu/zubal/



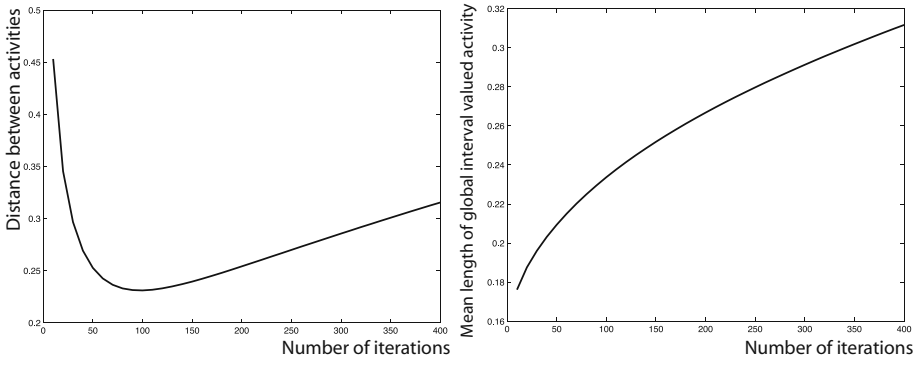
(a) Zubal phantom (b) 70 iterations (c) 400 iterations

Fig. 1. (a) The Zubal phantom and two reconstructed median images for 70 iterations (b) and 400 iterations (c)

realization, the activity is reconstructed with the NIBART method for different number of iterations. We account for the knowledge on the noise in the projections by considering $[\mathbf{S}]$ as being the 99% confidence interval under the hypothesis of a Poisson noise. The relaxation parameter has been empirically set to $\lambda = 0.024$ to ensure the convergence of the algorithm. Let $[\mathbf{I}^{s,t}] = [\underline{\mathbf{I}}^{s,t}, \overline{\mathbf{I}}^{s,t}]$ $(s \in \{1, \ldots, 100\}, t \in \{10, 20, \ldots, 400\})$ be the interval valued NIBART reconstruction of the s^{th} sinogram realization with t iterations. Let $\tilde{\mathbf{I}}^{s,t} = \frac{1}{2}(\underline{\mathbf{I}}^{s,t} + \overline{\mathbf{I}}^{s,t})$ be the median of $[\mathbf{I}^{s,t}]$. Let $\Delta\mathbf{I}^{s,t} = \overline{\mathbf{I}}^{s,t} - \underline{\mathbf{I}}^{s,t}$ be the spread of $[\mathbf{I}^{s,t}]$. For each pixel n $(n \in \{1, \ldots, N\})$ we compute $\delta_n^t = \frac{1}{100} \sum_{s=1}^{100} \Delta I_n^{s,t}$ the mean spread of the n^{th} pixel's interval valued activity reconstructed with t iterations, and σ_n^t the standard deviation of the median values $(\tilde{I}_n^{s,t})_{s \in \{1,\ldots,100\}}$. This experiment being simulated, the convergence of the reconstructed image to the original image through the iterations can be easily depicted by computing the Euclidian distance between those two images. This distance is plotted on Fig.2(a) for one realization. For all experiments, the smallest Euclidian distance value is obtained for $t \approx 70$. We thus will consider that, after 70 iterations, the algorithm has converged. The weak adequacy occurs after always before the real convergence. In this experiment, the weak adequacy is obtained as soon as $t \geq 6$

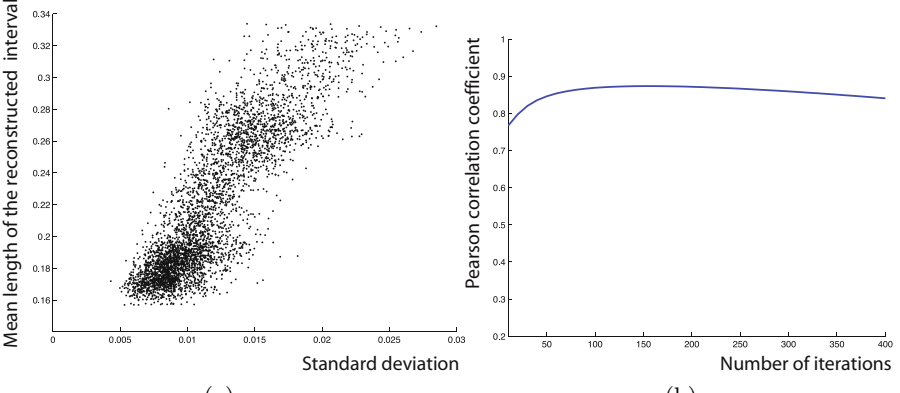
Fig.1 presents one of the reconstructed median images for 70 and 400, that is for convergence and far from convergence. This figure shows that, since the NIBART algorithm is not regularized, the spurious features increase with the iterations.

We aim at testing wether $\delta^t_{n \in \{1,\ldots,N\}}$ can be considered as a measure of the noise level of the activity reconstructed activity. This is achieved by testing wether the distributions $\sigma^t_{n \in \{1,\ldots,N\}}$ and $\delta^t_{n \in \{1,\ldots,N\}}$ are correlated whatever the number t of iterations. Fig.3(a) plots the clouds of $\sigma^{70}_{n \in \{1,\ldots,N\}}$ versus $\delta^{70}_{n \in \{1,\ldots,N\}}$. It shows a high correlation between the statistical reconstruction error and its estimation by using the length of the reconstructed interval in convergence conditions. Fig.3(b) plots the linear (Pearson) correlation coefficient between $\sigma^t_{n \in \{1,\ldots,N\}}$ and $\delta^t_{n \in \{1,\ldots,N\}}$ for different values of t. Fig.3(b) highlight the fact that, as soon as the convergence is obtained (i.e. after 70 iterations), the



(a) (b)

Fig. 2. (a) Distance between the simulated activity and a reconstructed median image, (b) Evolution of the mean length of the global interval valued activity through the iterations



(a) (b)

Fig. 3. (a) The standard deviation of the reconstructed activity versus its estimation by using the length of the reconstructed intervals for the 70^{th} iterations, (b) Evolution of the correlation coefficient through the iterations

standard deviation of the median value is highly correlated with the length of the reconstructed interval. Moreover, this correlation remains high even when the reconstructed image is far from convergence. In fact, when the convergence conditions are not fulfilled, the average length of the interval-valued reconstructed activity tends to increase (see Fig. 2(b)), reflecting the fact that the influence of the noise in the measurement on the reconstructed image increases with the iterations. Note that, when the weak adequacy occurs, the correlation is already high (here ≥ 0.77).

7 Conclusion

In this paper, we have presented an interval based algebraic method for predicting the noise in emission tomography reconstructed images. This method is based on extending the SIRT method to account for the new modeling of the projection performed by the tomographic device presented in [5]. The main characteristic of our approach is that the reconstructed activity is interval valued. As highlighted in the simulated experiment we report here, the length of the interval associated to each pixel of the reconstructed image is highly correlated with the statistical variance of its reconstructed value. Such a noise prediction can be necessary for making reliable the comparison between the reconstructed activity in two regions of interest, this comparison being the ground of nuclear imaging diagnosis. In future work, we will consider extending this method to other iterative reconstruction techniques like EM algorithm [12]. We will carry out experiments with real data obtained with a physical phantom. We will also compare our approach to the method proposed by Stayman and Fessler [2] on both simulated and real data.

References

1. Barret, H., Wilson, D., Tsui, B.: Noise properties of the EM algorithm. I: theory. Phys. Med. Bio. 39, 833–846 (1994)
2. Stayman, J., Fessler, J.: Efficient calculation of resolution and covariance for penalized-likelihood reconstruction in fully-3D SPECT. IEEE Trans. on Image Processing 23, 1543–1556 (2004)
3. Qi, J., Leahy, R.M.: A theoretical study of the contrast recovery and varaince of MAP reconstructions from PET data. IEEE Trans. Med. Imag. 18(4), 293–305 (1999)
4. Budinger, T., Derenzo, S., Greenberg, W., Gullberg, G., Huesman, R.: Quantitative potentials of dynamic emission computed tomography. J. Nucl. Med. 19, 309–315 (1978)
5. Rico, A., Strauss, O., Mariano-Goulart, D.: Choquet integrals as projection operators for quantified tomographic reconstruction. Fuzzy Sets and Systems 160, 198–211 (2009)
6. Gordon, R., Bender, R., Herman, G.: Algebraic reconstruction techniques (ART) for three dimensional electron microscopy and x-ray photography. J. Theor. Bio. 36, 105–117 (1970)
7. Dempster, A., Laird, N., Rubin, D.: Maximum likelihood from incomplete data via the EM algorithm. J. Royal Stat. Soc. 39, 1–38 (1977)
8. Hudson, H., Larkin, R.: Accelerated image reconstruction using ordered subsets of projection data. IEEE Trans. Med. Imag. 13, 601–609 (1994)
9. Lakshminarayanan, A., Lent, A.: Methods of least squares and SIRT in reconstruction. J. Theor. Bio. 76, 267–295 (1979)
10. Mariano-Goulart, D., Maréchal, P., Gratton, S., Giraud, L., Fourcade, M.: A priori selection of the regularization parameters in emission tomography by Fourier synthesis. Computerized Medical Imaging and Graphics 31, 502–509 (2007)
11. Danilov, V., Koshevoy, G.: Methods of least squares and SIRT in reconstruction. J. of Math. Anal. and App. 247, 1–14 (2000)
12. Hebert, T., Leahy, R.: A generalized EM algorithm for 3-D bayesian reconstruction from Poisson data using gibbs priors. IEEE Trans. Med. Imag. 8, 194–202 (1989)

A Log-Euclidean Polyaffine Registration for Articulated Structures in Medical Images

Miguel Ángel Martín-Fernández, Marcos Martín-Fernández, and Carlos Alberola-López

Laboratory of Image Processing, ETSI Telecomunicaciones, University of Valladolid, Campus Miguel Delibes s/n, 47011, Valladolid Spain
{migmar,marcma,caralb}@tel.uva.es

Abstract. In this paper we generalize the Log-Euclidean polyaffine registration framework of Arsigny et al. [1] to deal with articulated structures. This framework has very useful properties as it guarantees the invertibility of smooth geometric transformations. In articulated registration a skeleton model is defined for rigid structures such as bones. The final transformation is affine for the bones and elastic for other tissues in the image. We extend the Arsigny el al.'s method to deal with locally-affine registration of pairs of wires. This enables the possibility of using this registration framework to deal with articulated structures. In this context, the design of the weighting functions, which merge the affine transformations defined for each pair of wires, has a great impact not only on the final result of the registration algorithm, but also on the invertibility of the global elastic transformation. Several experiments, using both synthetic images and hand radiographs, are also presented.

1 Introduction

Image registration is the determination of a geometric transformation that maps one image into another, aligning objects in both images. Registration within medical imaging is of particular interest [2]. Methods proposed within this field have been classified as frame-based, landmark-based, surface-based, and intensity-based. With respect to the geometrical transformation, methods can be either rigid (only rotations and translations) [3], non-rigid (including scaling, affine transformations, projections and perspective) [4] and elastic transformations [5]. The latter are widely used since local features can be matched while continuity and smoothness in the transformation is maintained.

Very few methods have been proposed in the literature to register images including articulated structures. That is the case for medical images including bones for which rigid transformations are required, but also including surrounding soft tissues for which elastic registration should be used. Papademetris et al. [6] combine several rigid transformations of joints for lower-limb mouse images, obtaining a continuous overall transformation. Baiker et al. [7] propose

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 156–164, 2009.

a 3D hierarchical anatomical model in mice skeletal systems. Li et al. [8] propose an algorithm to register whole-body inter-subject volumes for mice and humans. They register the bone structures and then this transformation is refined for the other tissues. du Bois d'Aische et al. [9], present an articulated registration method for the spine column, based on the finite elements method for the elastic deformation outside rigid structures. No invertibility study for these methods, nor an evaluation of the sensitivity of the results with respect to the weighting functions used to fuse the set of rigid transformations are presented.

Arsigny et al. [10] propose a method in which local rigid or affine transformations are combined in a way that the overall transformation is invertible. This is a very important property of the searched transformation. They applied their method to the registration of histological images, arguing that it could also be used for articulated structures. In [1], the authors propose a Log-Euclidean polyaffine transformation approach where its inverse is always guaranteed, being this inverse another Log-Euclidean polyaffine transformation. The authors present a method to merge several affine transformations using normalized weighting functions. Their work focuses on the determination of the global transformation with special stationary properties which provides its invertibility. In [11], the authors proposed Gaussian weights for the regions of influence, where each weight is obtained by means of a mixture of Gaussians placed at some anchor points. In [4], a general framework for articulated registration of images with an inner bonny structure which guarantees exact registration in bone axes while allowing smooth transformation of soft tissues is presented. We propose an iterative process to obtain weights guaranteeing this condition, but no study of the invertibility of overall transformation was performed.

In this paper we generalize the Arsigny et al. [1] Log-Euclidean polyaffine transformation scheme. This scheme guarantees the invertibility of the transformation. However, it cannot be applied directly to articulated registration problems as the affine transformations have to be defined in pairs of landmarks. In the case of articulated registration, each rigid structure is modeled as a wire, and the affine transformation should be applied to pairs of wires. Thus, we propose a generalization of this Log-Euclidean polyaffine transformation to deal with affine wire transformations. This new framework is able to tackle articulated structures such as the bone age assessment problem presented in [4]. We also evaluate the influence of the weighting functions, needed to merge the affine transformations, both in the registration results and in the invertibility of the transformation.

The paper is structured as follows. In section 2 we review articulated registration as well as Log-Euclidean polyaffine registration. We also present novel weighting functions to deal with articulated structures within this polyaffine framework. In section 3 we present several experiments, making a brief discussion about the achieved results. Finally, in section 4 we give some conclusions.

2 Method

2.1 Articulated Registration

Let us briefly recall the articulated registration framework, described in [4]. The method is a landmark-based elastic registration procedure for anatomical structures that bears an inner skeleton, such as the hand bones. The inner bone skeleton is modeled with a wire model, where wires are drawn by connecting landmarks located in the main joints of the skeletal structure to be registered (long bones). The main feature of this registration method is that it allows to perform affine and elastic transformations on the same image. In this way, in points corresponding to the bone axes (specifically, where the wires are located) an exact registration is guaranteed, while for the remaining image points an elastic registration is carried out.

Fig. 1(a) and Fig. 1(b) show landmarks superimposed on two radiographs. A wire model is built by joining with straight segments (rods) every pair of consecutive landmarks in each finger, following bones axes, as shown in Fig. 1(c).

After obtaining an affine transformation matrix $\mathbf{M_i}$ for each rod, R_i, in the model, in order to calculate the final articulated transformation, a weighting of all the transformations is made, with $w_i(x, y)$ a measure of how the transformation of R_i influences pixel (x, y) in the image. These weights are function of the distance of each image pixel to the rod. This operation can be expressed in homogeneous coordinates as

$$(u\ v\ 1)^T = \left[\sum_{i=1}^{N} w_i(x,y)\mathbf{M_i}\right](x\ y\ 1)^T,$$

where (u, v) and (x, y) are the Cartesian coordinates of each pixel in the moving image before (input space) and after (output space) the transformation.

To obtain the elastic —articulated— transformation it must satisfy the following conditions [4] (numbered for future reference). (1) the transformation must be affine for pixels nearby the wires, therefore in pixels on rod R_i, the weight $w_i(x, y)$ corresponding to the i-th rod transformation must have a value

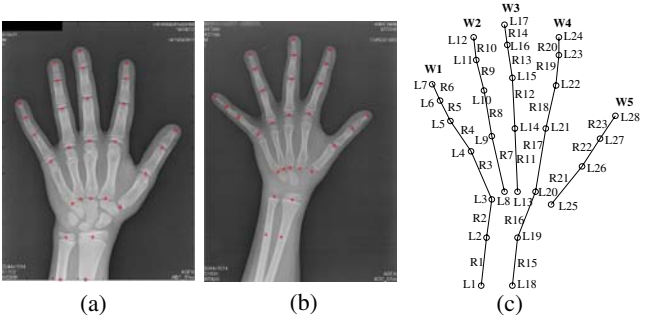


Fig. 1. (a) Landmark positions superimposed on a radiograph (used as fixed in the experiments); (b) Scheme of the wire model; (c) One example of moving image

near 1. (2) in pixels on rod R_i the other weights, $w_j(x, y)$, $j \neq i$, must have a value near 0. (3) in order to obtain a smooth transformation in points (x, y) far from the rods, the weight $w_i(x, y)$ should be defined in inverse proportion with the distance $d_i(x, y)$ of that pixel to rod R_i. (4) the sum of all the weights in each pixel (x, y) must be unity. To satisfy this last condition, it is enough to normalize, pixelwise, each weight with the sum of all of them.

2.2 Log-Euclidean Polyaffine Registration

In [1], the authors extend some previous works [10,11] dealing with polyaffine transformation, a mixture of locally affine transformation which fuses to a globally elastic transformation, proposing a Log-Euclidean polyaffine framework for registration. In particular, in [1], they proof that the transformation is invertible, being this inverse a new polyaffine transformation. In addition, they show that the transformation is invariant with respect to coordinate system. This two properties are very desirable in general for the registration of medical images.

The Log-Euclidean polyaffine transformation is proposed as the solution of the ordinary differential equation

$$\frac{d}{dt}(x\ y\ 1)^T = V(x, y) = \sum_{i=1}^{N} w_i(x, y) \log(\mathbf{M_i})\ (x\ y\ 1)^T.$$

The solution of this equation is always well-defined. The transformed value is given by integrating this equation between 0 and 1 with initial condition the starting point. In case the weights are constant, the solution is $\exp \sum_i \log(\mathbf{M_i})$, i.e. the Log-Euclidean mean of the components. The solution of this equation has very remarkable properties [1]. The implementation is given by means of the fast polyaffine transform. The speed $V(x)$ is scaled by 2^N (infinitesimal deformation), then a exponentiation is carried out (approximation) and the solution is given by means of N squarings (N compositions of the infinitesimal solution). The solution of the infinitesimal equation is approximated by

$$\mathbf{T}(x, y) = \sum_{i=1}^{N} w_i(x, y) \exp\left(\frac{1}{2^N} \log(\mathbf{M_i})\right)$$

and then N compositions (squarings) of this transformation is the approximate solution for the Log-Euclidean polyaffine transformation.

2.3 Weights Definitions

We propose different weight definitions suited for articulated registration that can be used within the Log-Euclidean framework presented in section 2.2. All the weighting functions depend on the distance $d_i(x, y)$ of each pixel (x, y) to each rod R_i. We recall to conditions (1)-(4) defined in section 2.1.

Arsigny et al.[11] modeled weights as a mixture of Gaussians located at several anchor points considered as representative of each region. In our case, the representative points of each region are the points on the rod, so we have generalized the Gaussian model to depend on the distance map of each rod, $d_i(x, y)$

with mean 0 and standard deviation σ_i, so that weight is maximum in rod pixels, and decays with distance to it. These weights are defined as

$$w_i^a(x,y) = G_{(0,\sigma_i)}(d_i(x,y)),$$

where $G_{(a,\sigma)}$ is a Gaussian function with mean a and standard deviation σ. These weights satisfy conditions (3) and (4) after normalization, but conditions (1) and (2) are not satisfied as a Gaussian is only zero at an infinite distance.

In [12] we proposed a set of weights. These weights, after an iterative process, satisfy approximately all the conditions (1)-(4). Initial weights $w_{i,0}^b(x,y)$ are calculated as a distance transform, with a value of 1 on the rod and 0 at an infinite distance from it, by means of

$$w_{i,0}^b(x,y) = \mathrm{sigm}\left(\frac{1}{1+d_i(x,y)}\right),$$

where $\mathrm{sigm}(\cdot)$ is a sigmoid function, which smooths the slope of the weight map near the wire. Finally weights are normalize to satisfy condition (4) using $w_{i,M}^b(x,y)$, where M is the total number of iteration steps.

We present here three new weight models that satisfy conditions (1)-(4) by definition, without the need of an iterative process

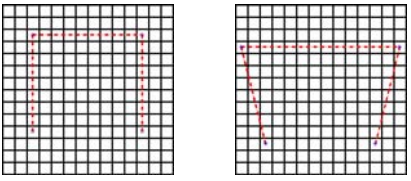
$$w_i^c(x,y) = \exp\left(-\frac{d_i(x,y)}{\prod_{j\neq i} d_j(x,y)^{\frac{1}{N-1}}}\right),$$

$$w_i^d(x,y) = \exp\left(-\frac{d_i(x,y)}{\min_{j\neq i} d_j(x,y)}\right), \qquad w_i^e(x,y) = G_{(0,1)}\left(\frac{d_i(x,y)}{\min_{j\neq i} d_j(x,y)}\right),$$

with N the number of rods. These three weights, w_i^c, w_i^d, w_i^e, are inversely proportional to $d_i(x,y)$ (distance from its rod) and directly proportional to the rest of distances. They have the advantage over w_i^b, that they do not need the iterative process to be, in these cases, exactly 1 on rod R_i, and 0 on the other rods, and they do not need any parameter, while w_i^b and w_i^a do. Comparing w_i^c and w_i^d, the first one is proportional to the geometric mean of the rest of distances, while the second one is proportional only to the dominant one, i.e. the minimum one, which has been proven to be more robust. Finally, w_i^e has lower gradient for pixels near the rod in the same way as sigmoid function in weight w_i^b.

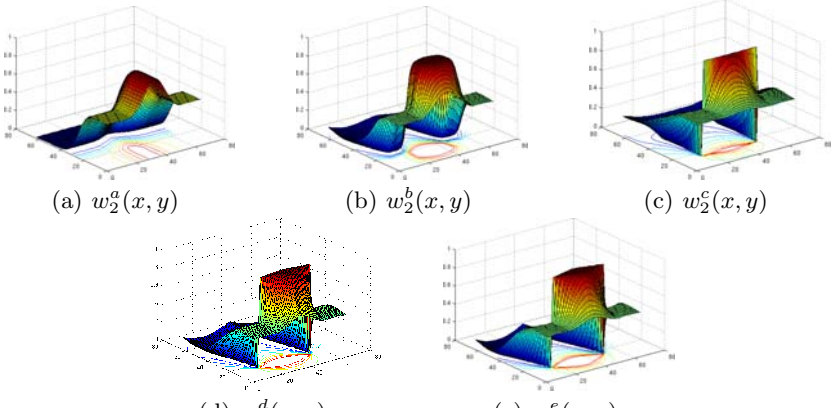
3 Results and Discussion

We have applied the articulated registration with the Log-Euclidean polyaffine framework to the synthetic images shown in Fig. 2, using all the weights defined in section 2.3. The weight maps for the horizontal rod, R_2, are shown in Fig. 3. It can be observed that w_2^a does not satisfy conditions (1) and (2), and w_2^b satisfies them approximately, while the other three ones satisfy them exactly. After normalization, weights w_2^d and w_2^e are very similar.



(a) Fixed image. (b) Moving image.

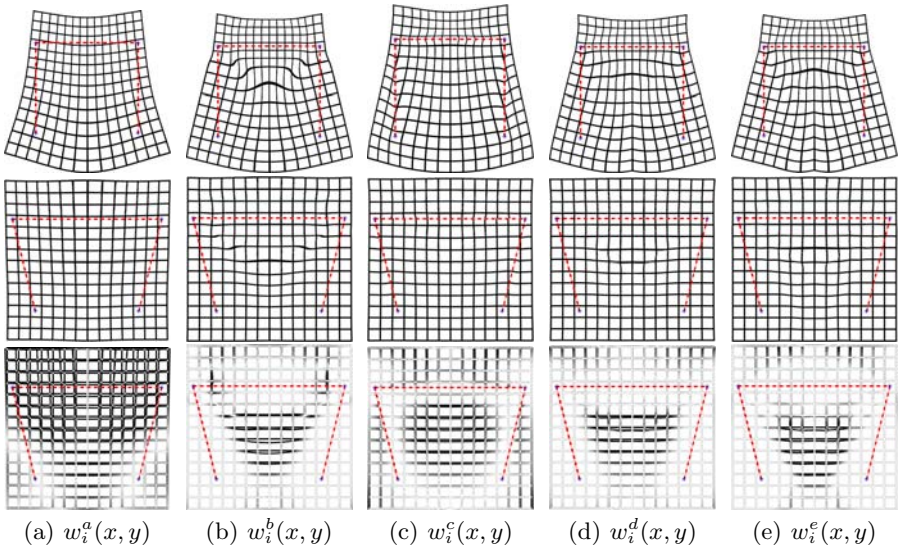
Fig. 2. Synthetic images with wire models superimposed in red and landmarks in blue



(a) $w_2^a(x,y)$ (b) $w_2^b(x,y)$ (c) $w_2^c(x,y)$

(d) $w_2^d(x,y)$ (e) $w_2^e(x,y)$

Fig. 3. Normalized weight maps corresponding to the horizontal rod R_2 in Fig. 2(a)

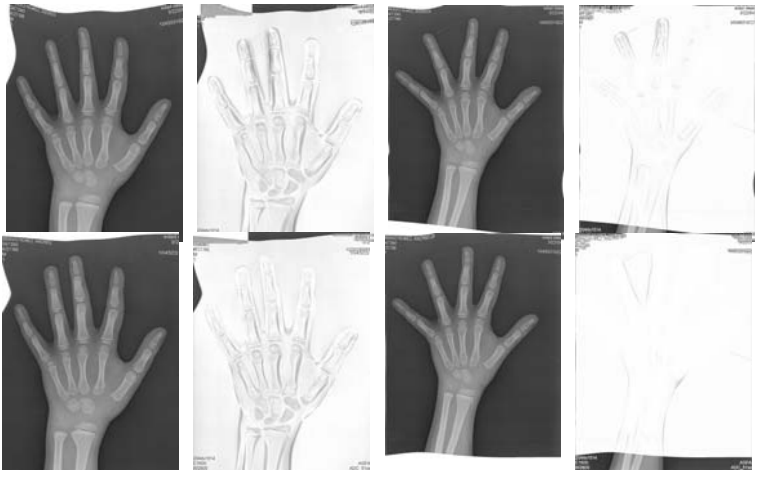


(a) $w_i^a(x,y)$ (b) $w_i^b(x,y)$ (c) $w_i^c(x,y)$ (d) $w_i^d(x,y)$ (e) $w_i^e(x,y)$

Fig. 4. Direct transformation of moving image (upper row), inverse transformation (middle) and difference of inverse transformation with fixed image for images of Fig. 2

Table 1. Mean and standard deviation of the Mutual Information obtained from 5 radiographs for direct and inverse transformation, when using the different weights

	Direct		Inverse	
Weight	**Mean**	**Std dev.**	**Mean**	**Std dev.**
$w_i^a(x,y)$	1.32	0.11	2.65	0.32
$w_i^b(x,y)$	1.34	0.08	2.68	0.30
$w_i^c(x,y)$	1.06	0.10	2.96	0.21
$w_i^d(x,y)$	1.47	0.07	3.27	0.38
$w_i^e(x,y)$	1.49	0.07	3.41	0.39



(a) Direct (b) Direct error (c) Inverse (d) Inverse error

Fig. 5. Graphical registration results for $w_i^a(x,y)$ (upper row) and $w_i^e(x,y)$ (lower row)

Transformed and inverse transformed moving image are shown in Fig. 4 together with the difference between moving and inverse transformed image. For the Log-Euclidean polyaffine algorithm $N = 6$ squaring steps are employed. It can be seen that w_i^a does not guarantee exact transformations of pixels near the rods, while all the other ones do. w_i^b produces the highest deformation in the middle of the rods, causing some residual errors, and w_i^c produces the most abrupt direct transform. w_i^d and w_i^e give similar results.

Finally, we have applied the method to five different hand radiographs which have been registered using image shown in Fig.1(a) as fixed image. $N = 6$ squaring steps are also employed. Table 1 shows numerical results of Mutual Information (MI) obtained for direct and inverse registration. Fig. 5 shows the graphical results when Fig.1(c) is used as moving image, with w_i^a and w_i^e, where it can be seen misregistration produced by w_i^a. With regard to the direct registration, w_i^e and w_i^d show the best performance in terms of MI, while w_i^c shows the worst results. Regarding invertibility of the transformation, w_i^e shows the best performance, while w_i^a is the worst one. Graphical results confirm that registration with w_i^e beats all the others in both directions.

4 Conclusions

We have generalized the Log-Euclidean polyaffine registration framework to articulated structures and compared different weight definitions. From the results obtained we can conclude that Gaussian weights, w_i^a, do not guarantee the rigidity and accuracy of the transformations in the bonny skeleton, while all the others do. Numerical and graphical results confirm that w_i^e outperforms the other weights in direct and inverse transformations.

Acknowledgments. The authors acknowledge the Spanish CICYT for research grant TEC 2007-67073/TCM. Thanks also go to Dr. Andrés de Llano and Dr. B. Viñuela from Hospital Río Carrión, Palencia, Spain, and Dr. S. Alberola, Centro de Salud Jardinillos, Palencia, Spain, for their valuable comments of the medical aspects of this work, and for the radiographs collected in the Radiology Department of the above mentioned hospital for research purposes.

References

1. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A Log-Euclidean Polyaffine Framework for Locally Rigid or Affine Registration. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 120–127. Springer, Heidelberg (2006)
2. Maintz, J.B.A., Viergever, M.A.: A Survey of Medical Image Registration. Med. Image Anal. 2, 1–36 (1998)
3. Maes, F., Collingnon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality Image Registration by Maximization of Mutual Information. IEEE Trans. Med. Imaging 16, 187–198 (1997)
4. Martín-Fernández, M.A., Cárdenes, R., Muñoz-Moreno, E., de Luis-García, R., Martín-Fernández, M., Alberola-López, C.: Automatic Articulated Registration of Hand Radiographs. Image Vis. Comput. (2008) doi:10.1016/j.imavis.2008.11.001
5. Rohr, K., Stiehl, H., Buzug, T., Weese, J., Kuhn, M.: Landmark-Based Elastic Registration Using Approximating Thin-Plate Splines. IEEE Trans. Med. Imaging 20, 526–534 (2001)
6. Papademetris, X., Dione, D.P., Dobrucki, L.W., Staib, L.H., Sinusas, A.J.: Articulated Rigid Registration for Serial Lower-Limb Mouse Imaging. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 919–926. Springer, Heidelberg (2005)
7. Baiker, M., Milles, J., Vossepoel, A., Que, I., Kaijzel, E., Löwik, C., Reiber, J.H.C., Dijkstra, J., Lelieveldt, B.: Fully Automated Whole-Body Registration in Mice Using an Articulated Skeleton Atlas. In: IEEE ISBI, pp. 728–731 (2007)
8. Li, X., Peterson, T.E., Gore, J.C., Dawant, B.M.: Automatic Registration of Whole Body Serial Micro CT Images with a Combination of Point-Based and Intensity-Based Registration Techniques. In: IEEE ISBI, pp. 18–25 (2006)
9. du Bois d'Aische, A., Craene, M.D., Geets, X., Gregoire, V., Macq, B., Warfield, S.K.: Estimation of the Deformations Induced by Articulated Bodies: Registration of the Spinal Column. Biomed. Signal Process. and Control 2, 16–24 (2007)

10. Arsigny, V., Pennec, X., Ayache, N.: Polyrigid and Polyaffine Transformations: A New Class of Diffeomorphisms for Locally Rigid or Affine Registration. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2879, pp. 829–837. Springer, Heidelberg (2003)
11. Arsigny, V., Pennec, X., Ayache, N.: Polyrigid and Polyaffine Transformations: a Novel Geometrical Tool to Deal with Non-Rigid Deformations - Application to the Registration of Histological Slices. Med. Image Anal. 9, 507–523 (2005)
12. Martín-Fernández, M.A., Muñoz Moreno, E., Martín-Fernández, M., Alberola-López, C.: Articulated Registration: Elastic Registration Based on a Wire-Model. In: Fitzpatrick, J.M., Reinhardt, J.M. (eds.) Medical Imaging 2005: Image Processing. SPIE Press, San Diego (2005)

Nonrigid Registration of Myocardial Perfusion MRI Using Pseudo Ground Truth

Chao Li and Ying Sun

Department of Electrical and Computer Engineering,
National University of Singapore, Singapore 117576, Republic of Singapore
{chaoli,elesuny}@nus.edu.sg

Abstract. In this paper we present a method for nonrigid registration of myocardial perfusion MR images. Instead of registering pairs of images within the observed sequence, we register the observed sequence to a pseudo ground truth, which is a motion/noise-free sequence estimated from the observed one. As the corresponding images of the two sequences have almost identical intensity distributions, our method overcomes the challenges arising from rapidly varying image intensity and contrast. The pseudo ground truth and the deformation fields for the observed sequence are obtained simultaneously by minimizing an energy functional integrating both the registration error and the spatiotemporal constraints on the pseudo ground truth in an expectation-maximization fashion. We have tested the proposed nonrigid registration method on real cardiac MR perfusion scans, both qualitatively and quantitatively. Experimental results show that the proposed method is able to successfully compensate for the heart motion during contrast enhancement.

1 Introduction

Myocardial perfusion magnetic resonance imaging (MRI) has demonstrated great potential for diagnosing cardiovascular diseases. In a myocardial perfusion MRI study, the heart is scanned along short-axis slices repeatedly at the same phase of the cardiac cycle through electrocardiographic gating, following a bolus injection of a contrast agent. Patient breathing during image acquisition often causes large variations in the position of the left ventricle (LV) in different frames. At the same time, the shape of the heart may change during contrast enhancement due to variations in heart rate. Therefore, it is desirable to perform nonrigid registration on time-series perfusion images to account for local elastic deformation.

The challenges in perfusion image registration mainly arise from rapid intensity changes of the heart ventricles during the wash-in/wash-out of the contrast agent. Mutual information (MI) [1] and normalized gradient fields [2,3] have been used to account for time-varying intensity. However, these methods are either computationally expensive or pure gradient based.

To minimize the intensity difference between the reference and floating images, some approaches achieve registration of an image sequence by registering every two consecutive frames [4,2]. However, this serial registration scheme tends to

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 165–172, 2009.

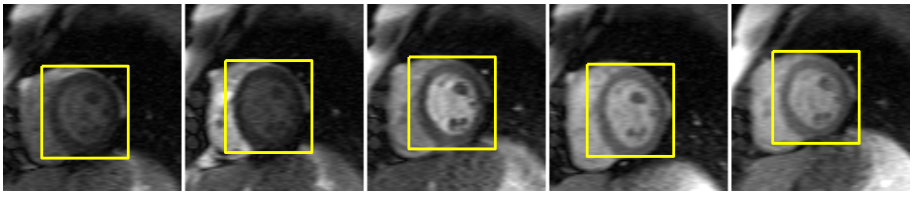


Fig. 1. Initial alignment of a myocardial perfusion MR image sequence

accumulate registration errors. In [3], Wollny et al. proposed to select the reference frame according to the breathing periodicity, but this method still requires registration of images from different perfusion phases. To avoid this problem, changes in image intensity were modeled by a Cluster-aware Active Appearance Model built from an annotated training set [5]. In another approach, Independent Component Analysis was used to generate a time-varying reference image for motion compensation [6]. However, in the above two approaches, it is not easy to choose the appropriate number of clusters or independent components.

In this paper, we propose a novel registration method that overcomes the challenges arising from time-varying intensity by using pseudo ground truth. The pseudo ground truth is an estimate of the image sequence that would have been acquired without being affected by motion or noise. We design an energy function that integrates both nonrigid registration and pseudo ground truth estimation, which can be minimized iteratively by solving a system of linear equations and applying nonrigid registration to corresponding pairs of images between the observed sequence and the pseudo ground truth sequence. In contrast to pairs of images within the observed sequence, these pairs of images have similar intensities, and the registration problem is therefore greatly simplified.

The rest of the paper is organized as follows. Section 2 describes our nonrigid registration algorithm. Section 3 presents the experimental results on real patient cardiac perfusion MRI data sets, followed by conclusion in Section 4.

2 Method

We propose to solve the nonrigid registration problem in three steps: 1) automatic selection of one reference frame and detection of a region of interest (ROI) that contains the LV; 2) initial alignment by identifying the global translation of the ROI in each frame; and 3) estimation of local elastic deformation. For the first two steps, we adopt the methods described in [7] and [8], respectively. In the following sections, we focus on the last step, i.e., nonrigid registration.

Fig. 1 displays the results after initial alignment for 5 selected frames from a myocardial perfusion MRI study, in which the bounding box of the ROI has been shifted to the best match location in each frame. By aligning the ROI from different frames, we obtain a roughly registered ROI sequence that still contains residual motion incurred by the local elastic deformation of the LV.

2.1 Energy Function for Nonrigid Registration

Instead of registering pairs of images within the image sequence, we propose to register each image to its counterpart in a pseudo ground truth image sequence that is assumed to be free of motion. This avoids the problem caused by unknown intensity variations between different frames, because registration is between images with almost identical intensity distributions.

Given an observed image sequence g, we solve for a pseudo ground truth sequence f and a nonrigid deformation function H that minimize the following energy functional:

$$E(H(g), f) = E_d(H(g), f) + \alpha E_s(f) + \beta E_t(f), \tag{1}$$

subject to the constraint on H that its underlying displacement field is zero for the reference frame, and is smooth for the rest frames. In (1), E_d is the data fidelity term that measures the difference between the pseudo ground truth sequence f and the nonrigidly deformed sequence $H(g)$; E_s is the spatial smoothness constraint penalizing the intensity difference between neighboring pixels of the same tissue type; and E_t is the temporal smoothness constraint penalizing the second order derivative of the intensity-time curve of each pixel; α and β are positive scalars that control the weights of different terms.

To rewrite the energy functional in matrix-vector form, each image sequence is represented as a column vector, e.g., if $g(i, j, t)$ is the intensity at pixel (i, j) in MRI frame t, then the column vector $\mathbf{g}$ is given by:

$$\begin{aligned}\mathbf{g} = \mathrm{vec}\,(g) = [\, & g(1,1,1) \ldots g(N_i,1,1) \ldots g(1,N_j,1) \ldots \\ & g(N_i,N_j,1) \ldots g(1,1,N_t) \ldots g(N_i,N_j,N_t)\,]^{\mathrm{T}},\end{aligned}$$

where N_i and N_j are respectively the number of rows and columns of each image, and N_t is the number of frames. Similarly, we define $\mathbf{f} = \mathrm{vec}(f)$ as the column vector of the pseudo ground truth f.

Data Fidelity Term. Let $\tilde{g} = H(g)$ denote the image sequence obtained by deforming g with H. The data fidelity term E_d is the sum-squared intensity difference between $\tilde{g}$ and f:

$$E_d(H(g), f) = E_d(\tilde{\mathbf{g}}, \mathbf{f}) = (\tilde{\mathbf{g}} - \mathbf{f})^{\mathrm{T}}(\tilde{\mathbf{g}} - \mathbf{f}), \quad \tilde{\mathbf{g}} = \mathrm{vec}(\tilde{g}) = \mathrm{vec}(H(g)). \tag{2}$$

Spatial Smoothness Constraint. Motivated by the observation that pixels of the same tissue type have similar intensities in each frame of the sequence, we impose a spatial smoothness constraint on f by penalizing the sum of weighted intensity differences between each pixel and its nearest neighbors:

$$E_s(f) = E_s(\mathbf{f}) = \sum_{k=1}^{K} (\mathbf{D}_k^s \mathbf{f})^{\mathrm{T}} \mathbf{W}_k (\mathbf{D}_k^s \mathbf{f}), \tag{3}$$

where K is the number of neighboring pixels being considered, and in this work we use $K = 4$; $\mathbf{D}_k^s$ is the first order spatial derivative operator along the direction

between each pixel and its kth neighbor, and $\mathbf{W}_k$ is the corresponding weight matrix derived from the intensity-time curves, such that intensity differences between neighboring pixels of different tissue types are not penalized.

Let $\mathbf{D}_f(N)$ and $\mathbf{D}_b(N)$ respectively denote the first order forward and backward derivative operators of size $N \times N$. The four derivative operators are then respectively given by $\mathbf{D}_1^s = \mathbf{I}_{N_t} \otimes \mathbf{I}_{N_j} \otimes \mathbf{D}_f(N_i)$, $\mathbf{D}_2^s = \mathbf{I}_{N_t} \otimes \mathbf{I}_{N_j} \otimes \mathbf{D}_b(N_i)$, $\mathbf{D}_3^s = \mathbf{I}_{N_t} \otimes \mathbf{D}_f(N_j) \otimes \mathbf{I}_{N_i}$, and $\mathbf{D}_4^s = \mathbf{I}_{N_t} \otimes \mathbf{D}_b(N_j) \otimes \mathbf{I}_{N_i}$, where $\otimes$ represents the Kronecker operator and $\mathbf{I}_N$ denotes the identify matrix of size $N \times N$.

Another observation is that pixels of the same tissue type have intensity-time curves that exhibit similar temporal dynamics. Therefore, we use the correlation coefficients between intensity-time curves of neighboring pixels to set the weight matrix $\mathbf{W}_k$. Here the intensity-time curves are obtained from the deformed image sequence $\tilde{g}$. Let ρ_{ijk} denote the correlation coefficient between the intensity vectors of pixel (i, j) and its kth neighbor, thus $\mathbf{W}_k = \mathrm{diag}(\mathrm{vec}(w_k))$ and

$$w_k(i,j,t) = \begin{cases} \rho_{ijk}, \text{ if } \rho_{ijk} > \rho_{th} \\ 0, \quad \text{otherwise} \end{cases} \tag{4}$$

where $\rho_{th} \in [0, 1)$ is a user-defined threshold to ensure that only intensity differences between neighboring pixels of the same tissue type are penalized.

Temporal Smoothness Constraint. As the contrast agent gradually perfuses through the heart, generally the intensity increase/decrease step-size does not vary much during the same perfusion phase. Therefore we impose a temporal smoothness constraint on f by penalizing the second order time derivative of the pseudo ground truth:

$$E_t(f) = E_t(\mathbf{f}) = (\mathbf{D}_2^t \mathbf{f})^{\mathrm{T}} (\mathbf{D}_2^t \mathbf{f}). \tag{5}$$

Let $\mathbf{D}_{tt}(N)$ denote the second order derivative operator of size $N \times N$. The second order time derivative operator is given by $\mathbf{D}_2^t = \mathbf{D}_{tt}(N_t) \otimes \mathbf{I}_{N_j} \otimes \mathbf{I}_{N_i}$.

2.2 Energy Minimization

In our implementation, the energy functional defined in (1) is minimized by iteratively solving for the optimal H and $\mathbf{f}$, in an expectation-maximization fashion. In each iteration, first we keep H fixed, and estimate the pseudo ground truth $\mathbf{f}$ by minimizing

$$E = (\tilde{\mathbf{g}} - \mathbf{f})^{\mathrm{T}} (\tilde{\mathbf{g}} - \mathbf{f}) + \alpha \sum_{k=1}^{K} (\mathbf{D}_k^s \mathbf{f})^{\mathrm{T}} \mathbf{W}_k (\mathbf{D}_k^s \mathbf{f}) + \beta (\mathbf{D}_2^t \mathbf{f})^{\mathrm{T}} (\mathbf{D}_2^t \mathbf{f}). \tag{6}$$

This minimization requires solving a system of linear equations:

$$\left[\mathbf{I} + \alpha \sum_{k=1}^{K} \left(\mathbf{D}_k^{s\,\mathrm{T}} \mathbf{W}_k \mathbf{D}_k^s \right) + \beta \mathbf{D}_2^{t\,\mathrm{T}} \mathbf{D}_2^t \right] \mathbf{f} = \tilde{\mathbf{g}},$$

which can be solved by using Gaussian elimination and results in an estimate $\hat{\mathbf{f}}$.

Then we perform registration between the reference image (static image) in g and its counterpart image (moving image) in $\hat{f}$ and the resulting deformation field is then applied to all the images in $\hat{f}$, which leads to a calibrated pseudo ground truth image sequence f so as to ensure that the underlying deformation field for the reference frame is zero. Next, we keep $\mathbf{f}$ fixed, and register each image in g to its counterpart. The resultant deformation fields completely and uniquely define the deformation function H, and hence $\tilde{\mathbf{g}}$ can be updated accordingly. Here we use the demons algorithm described in [9] to register corresponding images. Nevertheless, this pseudo ground truth based approach can be used in conjunction with many other registration methods.

The initial condition is $\tilde{\mathbf{g}} = \mathbf{g}$, i.e., the original observed image sequence, which implies the initial deformation function H is the identity. The weight matrices $\mathbf{W}_k, k = 1, \cdots, K$ are re-estimated using the updated $\tilde{\mathbf{g}}$ at each iteration. The iteration continues until E cannot be further reduced.

3 Experimental Results

We have tested the proposed method on 9 slices of perfusion images from 7 patients. The images were acquired on Siemens Sonata MR scanners following bolus injection of Gd-DTPA contrast agent. For all the data sets in our study, we consistently observed monotonic decrease of the energy function, which had little change after 3 or 4 iterations. Based on visual validation from video, the heart in the registered sequence was well stabilized. Next, we present the experimental results for a representative data set in two aspects: contour propagation and comparison of intensity-time curves.

3.1 Contour Propagation

Given the contour that delineates the boundary of the LV blood pool in the reference frame, one can propagate this contour to other frames according to the deformation field obtained using the method described in Section 2. Therefore, one way to evaluate the performance of the nonrigid registration algorithm is to verify whether the propagated contour well delineates the boundaries of the LV blood pool in other frames.

The first two rows in Fig. 2 compares the propagated contours before and after applying nonrigid registration for 4 consecutive frames from a real cardiac MR perfusion scan. These frames are chosen because the LV in these images undergoes noticeable shape changes. As shown in the top row, the contours before nonrigid registration do not lie exactly at the boundaries between the LV blood pool and the LV myocardium, especially at the regions that are indicated by the bright arrows. In contrast, the contours in the second row, after nonrigid registration using the proposed method, delineate well the boundaries between the LV blood pool and the LV myocardium.

To further demonstrate the performance of the proposed method, we compare our results with those obtained by serial registration using the demons algorithm [9], in which every two consecutive frames are registered to propagate

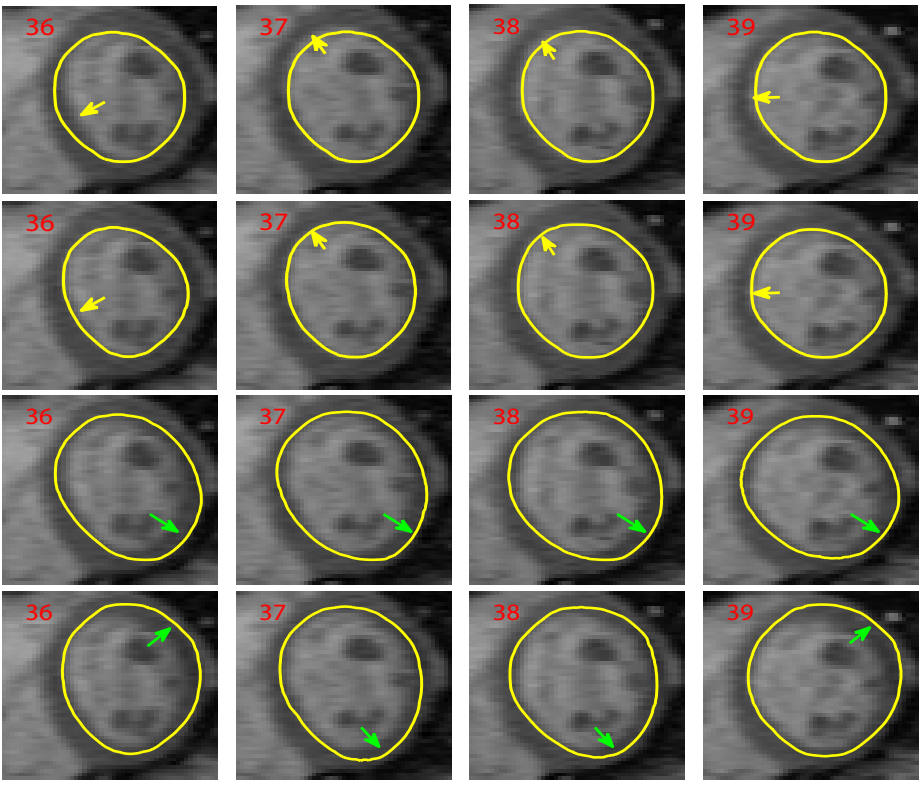


Fig. 2. Contour propagation for 4 consecutive frames from a real patient cardiac MR perfusion scan: contours before applying nonrigid registration (top row); contours propagated by our method (second row), serial demons registration (third row), and MI-based registration (bottom row)

from the reference frame to the rest of the sequence. As shown in the third row in Fig. 2, the propagated contours in all four frames are away from the true boundaries in the bottom right region due to the accumulation of registration errors. We have also applied the free-form deformation (FFD) registration method [10] using MI as the similarity measure, in which the reference frame is used to register all other frames in the sequence. As shown in the bottom row in Fig. 2, the propagated contours do not delineate the boundaries as accurately as those obtained by our method (see the second row in Fig. 2).

For quantitative evaluation, we manually drew the contours that delineate the LV blood pool in each frame for one slice, and measured the maximum and mean distances between the propagated contours and the corresponding manually-drawn contours. For the four frames shown in Fig. 2, the maximum distances were 2.9, 2.5, 1.7, and 1.7 pixels with only translation, as compared to 0.8, 0.9, 0.7 and 1.1 pixel(s) after applying our registration method. Over the entire sequence, the average mean distance was decreased from 0.70 pixel to 0.58

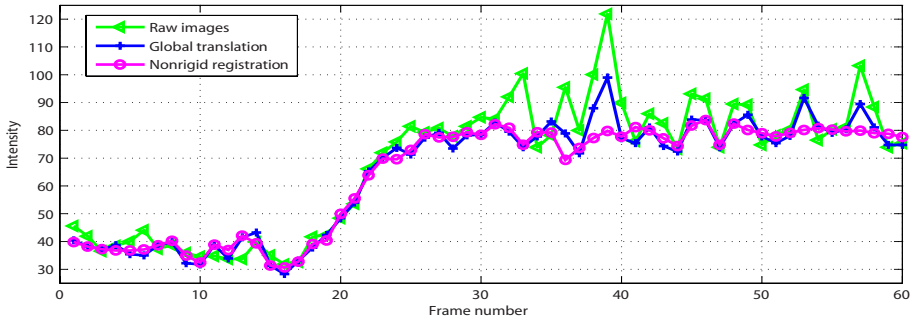


Fig. 3. Comparison of the average intensity-time curves of four neighboring pixels from the myocardium for the original (raw) data set, after compensating for global translation, and after performing nonrigid registration

pixel. In contrast, using the serial demons registration and the MI-based method, the average mean distance was increased to 1.40 and 1.95 pixels, respectively, due to accumulation of registration errors and large registration errors in frames having distinct intensity distributions from that of the reference frame.

3.2 Comparison of Intensity-Time Curves

In cardiac perfusion MRI, changes in pixel intensity corresponding to the same tissue across the image sequence are closely related to the concentration of the contrast agent. As the concentration of the contrast agent changes smoothly during the perfusion process, the intensity-time curves should be temporally smooth at pixels that are located on the LV myocardium.

We use a 2×2 pixel window to select a region in the LV myocardium, and the window is chosen to be close to the LV blood pool such that the motion of the LV can be reflected in the intensity-time curve of the window. Fig. 3 plots the average intensity-time curves within the window, for the original/raw data set (green triangle), after compensating for global translation (blue cross), and after performing nonrigid registration (magenta circle). For the original image sequence, the intensity-time curve exhibits very big oscillations in the second half of the sequence, in which the global translation is large due to patient breathing. After compensating for the global translation, the intensity-time curve becomes temporally smoother but still has some oscillations, e.g., in frames $37 - 40$ and $52 - 58$. This is because global translation is not capable of describing shape changes in these frames. However, after compensating for the local deformation by performing nonrigid registration, the intensity-time curve becomes smoother at frames when the LV undergoes noticeable local deformation, with only small local oscillations due to noise.

4 Conclusion

This paper presents a novel nonrigid registration algorithm for cardiac perfusion MR images. Unlike most registration methods that estimate the deformation

between pairs of images within the observed perfusion sequence, we introduce a pseudo ground truth to facilitate image registration. The pseudo ground truth is a motion-free sequence estimated from the observed perfusion data. Since the intensity distributions of the corresponding images between the pseudo ground truth sequence and the observed sequence are almost identical, this method successfully overcomes the challenges arising from intensity variations during perfusion. Our experimental results on real patient data have shown that our method is able to effectively compensate for the elastic deformation of the heart, and that it significantly outperforms the serial demons registration method and an MI-based method when registering myocardial perfusion images.

Acknowledgments. The authors thank Siemens Corporate Research, NJ, USA, for providing the datasets, and acknowledge the support by NUS grant R-263-000-470-112.

References

1. Ólafsdóttir, H.: Nonrigid registration of myocardial perfusion MRI. In: Proc. Svenska Symposium i Bildanalys (March 2005)
2. Wollny, G., Ledesma-Carbayo, M.J., Kellman, P., Santos, A.: A new similarity measure for non-rigid breathing motion compensation of myocardial perfusion MRI. In: Proc. IEEE EMBS, August 2008, pp. 3389–3392 (2008)
3. Wollny, G., Ledesma-Carbayo, M.J., Kellman, P., Santos, A.: Non-rigid motion compensation in free-breathing myocardial perfusion magnetic resonance imaging. In: Proc. Computers in Cardiology, September 2008, pp. 465–468 (2008)
4. Breeuwer, M., Spreeuwers, L., Quist, M.: Automatic quantitative analysis of cardiac MR perfusion images. In: Proc. SPIE, February 2001, pp. 733–742 (2001)
5. Stegmann, M.B., Larsson, H.B.W.: Motion-compensation of cardiac perfusion MRI using a statistical texture ensemble. In: Magnin, I.E., Montagnat, J., Clarysse, P., Nenonen, J., Katila, T. (eds.) FIMH 2003. LNCS, vol. 2674, pp. 151–161. Springer, Heidelberg (2003)
6. Milles, J., van der Geest, R.J., Jerosch-Herold, M., Reiber, J., Lelieveldt, B.: Fully automated motion correction in first-pass myocardial perfusion MR image sequences. IEEE Trans. Medical Imaging 27(11), 1611–1621 (2008)
7. Lorenz, C.H., Sun, Y., Meng, T., Luk, C., Kraitchman, D.L., Kirchberg, K.J.: Generation of MR myocardial perfusion maps without user interaction. In: Proc. Joint Annual Meeting ISMRM-ESMRMB 2007, Berlin, Germany (May 2007)
8. Sun, Y., Jolly, M.P., Moura, J.M.F.: Contrast-invariant registration of cardiac and renal MR perfusion images. In: Proc. 7th International Conference on Medical Image Computing and Computer Assisted Intervention, France (September 2004)
9. Wang, H., Dong, L., O'Daniel, J., Mohan, R., Garden, A.S., Ang, K.K., Kuban, D.A., Bonnen, M., Chang, J.Y., Cheung, R.: Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. Phys. Med. Biol. 50(12), 2887–2905 (2005)
10. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Trans. Medical Imaging 18(8), 712–721 (1999)

Parallax-Free Long Bone X-ray Image Stitching

Lejing Wang[1], Joerg Traub[1], Simon Weidert[2], Sandro Michael Heining[2], Ekkehard Euler[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
[2] Trauma Surgery Department, Klinikum Innenstadt, LMU Munich, Germany

Abstract. In this paper, we present a novel method to create parallax-free panoramic X-ray images of long bones during surgery by making the C-arm rotate around its X-ray source, relative to the patient's table. In order to ensure that the C-arm motion is a relative pure rotation around its X-ray source, we move the table to compensate for the translational part of the motion based on C-arm pose estimation, for which we employed a Camera Augmented Mobile C-arm system [1] and a visual planar marker pattern. Thus, we are able to produce a parallax-free panoramic X-ray image that preserves the property of linear perspective projection. We additionally implement a method to reduce the error caused by varying intrinsic parameters of C-arm X-ray imaging. The results show that our proposed method can generate a parallax-free panoramic X-ray image, independent of the configuration of bone structures and without the requirement of a fronto-parallel setup or any overlap in the X-ray images. The resulting panoramic images have a negligible difference (below 2 pixels) in the overlap between two consecutive individual X-ray images and have a high visual quality, which promises suitability for intra-operative clinical applications in orthopedic and trauma surgery.

1 Introduction and Related Works

Orthopedic and trauma surgery rely heavily on intra-operative X-ray images to visualize bone fracture configurations and to support repositioning and osteosynthesis, especially in minimally invasive surgery. Mobile C-arms are an everyday tool for acquiring X-ray images in the operating room during surgery. However, X-ray images acquired by mobile C-arms have a narrow field of view and often can not visualize the entire bone structure of interest within a single X-ray image, especially in long bone fixation surgery. Determining the relative position and orientation of bone segments across several individual images will definitely result in a compromise of accuracy and may make the surgery more complex. Panoramic X-ray images with exceptionally wide fields of view could be very helpful in these cases, since they can show the whole bone structure in a single image with a large field of view.

Panoramic X-ray images can be obtained by stitching many individual X-ray images. For intra-operative X-ray stitching one method was proposed by Yaniv and Joskowicz [2] using a standard mobile C-arm. This method introduces an orthopedic radiolucent X-ray ruler placed along and parallel to the bones. It

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 173–180, 2009.

uses the graduations of the ruler in the images to estimate the planar transformation by a feature-based alignment method and requires the user to manually select the reconstruction plane in order to compensate for parallax effects on that plane. However, this method requires overlapping areas between two consecutive X-ray images to estimate the planar transformation and thus needs additional radiation. Furthermore, it requires a fronto-parallel C-arm setup, i.e. the ruler plane must be parallel to the detector plane of the C-arm. Wang et al. [3] presented a method to generate panoramic X-ray images intra-operatively using a C-arm, to which a video camera was attached. This method does not require overlapping X-ray images or special X-ray markers. It uses the video images in combination with a visual marker pattern to estimate the planar transformation for creating panoramic X-ray images. In order to reduce the parallax effects on the bone plane, the marker plane and the bone plane need to be parallel to each other and the distance between them has to be manually estimated.

Parallax effects will cause ghosting and misalignment in the panoramas (see figure 1(a)). Although methods for reducing parallax effects on the bone plane were described in [2,3], they require the bone segments to be on the same plane and parallel to the plane used for estimating the transformation. In practice, it is unlikely that the plane of the bone is exactly parallel to the plane of the ruler (or marker pattern) due to the soft tissue around bones. Bone segments furthermore may be on different plane levels due to fracture or inability to fully extend the joints. For these complex situations, completely removing the parallax effects in the panoramic images is impossible by using the early introduced methods [2,3]. This will finally restrict the stitching approaches within clinical routine to a limited number of cases.

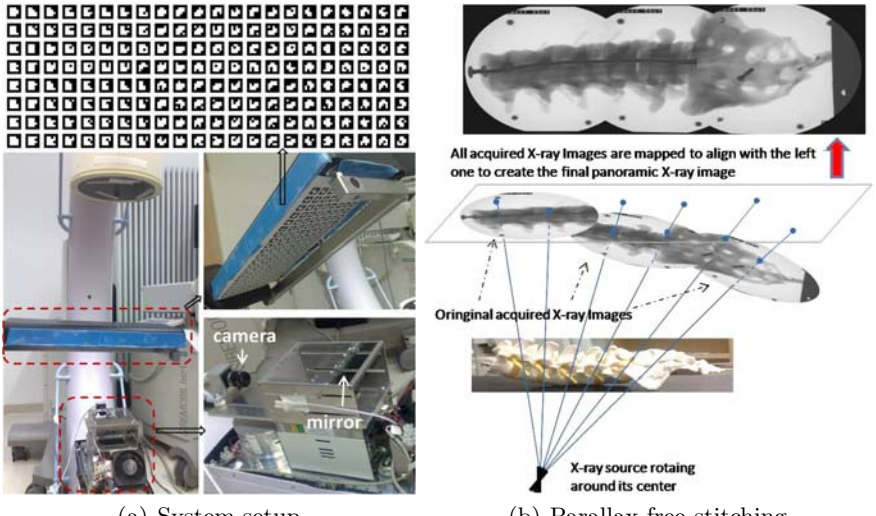
The pure rotation of the mobile C-arm around its X-ray source is a basic requirement for stitching individual X-ray images to a true parallax-free panoramic X-ray image, which preserves the linear perspective projection. However, making the mobile C-arm rotate around its X-ray source is impractical or even impossible, due to the limitation of the mechanical configuration of mobile C-arms. In this paper, we propose a new method for parallax-free X-ray image stitching by making the C-arm rotate around its X-ray source, relative to the patient's table. This is achieved by moving the table to compensate for the translational part of the motion based on C-arm pose estimation, for which we employed a Camera Augmented Mobile C-arm (CamC) system and a visual planar marker pattern. This CamC system that extends a standard mobile C-arm by a video camera and mirror construction was first proposed for X-ray and video image overlay [1]. Finally, we are able to generate a true parallax-free panoramic X-ray image, independent of the configuration of bone structures and without the requirement of a fronto-parallel setup or overlapping X-ray regions. This true panoramic X-ray image now can be treated as a single image obtained by a C-arm having an exceptional wide imaging field of view and also can be further processed by various computer vision algorithms that assume linear perspective projection.

(a) (b)

Fig. 1. (a) Illustration of the parallax effects in the panorama of stitching two X-ray images. From the left to right, the first image shows the construction of two metal keys on two different parallel planes. The second image is the panorama generated using the transformation for the plane of the smaller key, and the third image shows the panorama generated using the transformation for the plane of the bigger key. The last image presents a parallax-free panorama generated by the X-ray source undergoing a relative pure rotation; (b) Available movements for C-arms. 3 rotations around 3 orthogonal axes R_1, R_2 and R_3, 2 translations along 2 orthogonal axes T_1 and T_2.

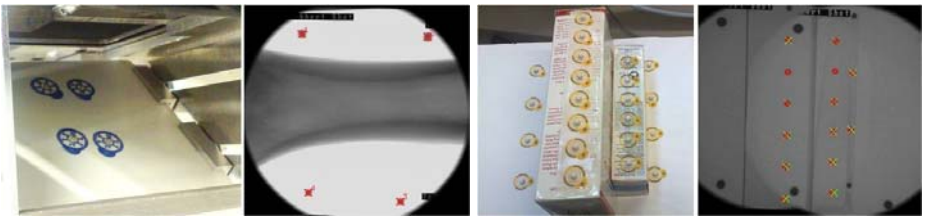
2 System Setup

Our system is composed of a planar square marker pattern, a translatable operating table and a CamC system built by attaching a video camera and mirror construction to a mobile C-arm. Our designed visual square marker pattern (see figure 2(a)) was printed in A2 size paper by a high definition printer. This marker



(a) System setup (b) Parallax-free stitching

Fig. 2. (a) The top image shows our square marker pattern. The bottom left image shows the setup for X-ray image stitching. The bottom right image shows the attachment of the marker pattern to the table back and our custom made camera and mirror construction; (b) A parallax-free panoramic image of a plastic lumbar spine and sacrum is generated by having the X-ray source rotate around its center relative to the object.



(a) VDP implementation (b) Parallax-free Stitching Quantification

Fig. 3. (a) The left image shows four X-ray markers attached on the back of the mirror. The right image shows the extracted centroids of the X-ray markers in the X-ray image marked by red crosses with indices; (b) The left image shows a phantom of spherical X-ray markers on three different planes. The right image shows that the extracted centroids of the X-ray markers are marked by red and green crosses in the two X-ray images stitched by using our method.

pattern was rigidly and flatly attached under the operating table. The mobile C-arm will be positioned in the common setup, in which the X-ray source is below the operating table (see figure 2(a)). In the method proposed by Wang et al. [3], the marker pattern was placed above the patient and could occlude the view onto the operation situs. Compared to their solution, our system setup does not only make the marker pattern visible to the camera for pose estimation, but also let the marker pattern be smoothly integrated into the surgical procedure without disturbing the physician's view. In order to reduce the error caused by varying intrinsic parameters of C-arm X-ray imaging, we attached four spherical X-ray markers with 1.5mm diameter on the back of the mirror, near the border of and within the X-ray's field of view (see figure 3(a)).

3 Method

The camera is attached to the C-arm such that its optical center virtually coincides with the X-ray source. A planar homography then is estimated for X-ray and video image overlay. This calibration enables that the X-ray source and the video camera have the same intrinsic and extrinsic parameters [1]. Therefore, all the poses and motions estimated using the video camera directly correspond to that of the X-ray projection geometry.

The key step of image stitching is the estimation of the planar homography to align images. The camera pose of acquiring the first image is defined as the relative world coordinate system, and the second image is obtained after a rotation $R \in \mathbb{R}^{3\times3}$ and a translation $t \in \mathbb{R}^3$ of the camera. The planar homography that aligns the first camera image to the second camera image is defined by $H = KRK^{-1} + \frac{1}{d}Ktn^TK^{-1}$ where $K \in \mathbb{R}^{3\times3}$ is the intrinsic matrix of the camera. H is valid for all image points whose corresponding space points are on the same plane, called stitching plane, defined by the normal vector $n \in \mathbb{R}^3$ and distance d to the origin in the world coordinate system [2]. However, any structure

that is not within this stitching plane in 3D space will get ghosting or blurring effects (see figure 1(a)) also refered to as parallax. H has two parts, KRK^{-1} and $\frac{1}{d}Ktn^TK^{-1}$, in which the second part relies on the plane parameters n and d. In case that the camera motion only contains the rotation R, $\frac{1}{d}Ktn^TK^{-1}$ is equal to zero. Then H is independent from the plane parameters, and thus it is valid for all image points without the coplanar constraint on their corresponding space points.

Parallax-Free X-ray Stitching. In real clinical cases, it is hard to assume that the bone structure is a simple and planar structure. Rotating the mobile C-arm around its X-ray source is an indispensable step for generating a panoramic X-ray image without any parallax effect. However, it is difficult and even impossible to make a pure rotation around the X-ray source, because standard mobile C-arms have only 5DOF (3 rotations around axes R_1, R_2 and R_3; 2 translations along axes T_1 and T_2) and there is no direct control of the X-ray source (see figure 1(b)). Therefore, we translate the patient's table in order to compensate for the translation of the X-ray source with respect to the table. The guidance is supported by pose estimation using the attached video camera. We designed a planar marker pattern, in which all the square markers can be uniquely detected [4]. The corners of each square marker can be extracted with subpixel accuracy and used as feature points. Having the marker pattern with known geometry, we are able to establish point correspondences between the 2D image points and 3D space points at the marker pattern. Based on these point correspondences, the pose of the camera (X-ray source) relative to the marker pattern (patient's table) is computed by using a camera calibration method [5]. The estimation of the pose is performed with a frequency of 8-10 Hz in our implementation. This can provide a continuous feedback for the translation of the table. Kainz et al. [6] developed a method to do C-arm pose estimation using an X-ray visible planar marker pattern attached to the operating table. It is not suitable as guidance for positioning of the table, since a continuous X-ray exposure and therefore a large amount of radiation is inevitable. After translation compensation, the rotation R of the camera can be computed from pose estimation. K was obtained during the CamC calibration. Therefore, the homography for aligning X-ray images can be computed and it is valid for all image points. To generate a parallax-free panoramic X-ray image, the first acquired X-ray image is defined as the reference image, and all others will be registered into its coordinate system (see figure 2(b)).

Model of C-arm X-ray Imaging with Fixed Intrinsic Parameters. The X-ray source and the detector plane are loosely coupled, mounted on the opposite sides of the C-arm. Due to gravity, it is hard to suppose a fixed relation between the X-ray source and the detector plane, i.e. intrinsic parameters change, in different C-arm orientations. However, the C-arm orientation must be changed to achieve a parallax-free stitching. In order to reduce the error caused by the changes of the C-arm orientation in the final panoramas, we adapted the idea

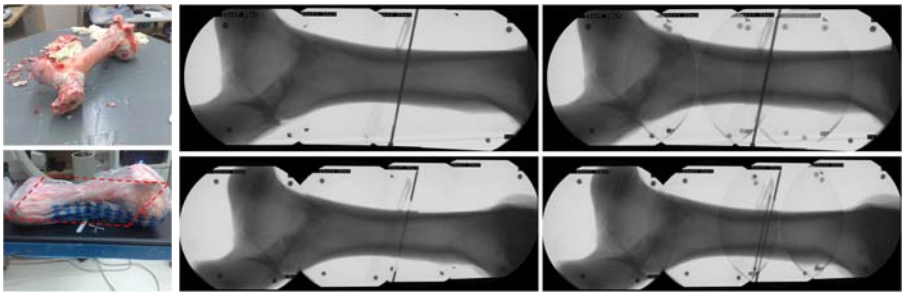
of the Virtual Detector Plane (VDP) proposed in [7]. Four X-ray markers, attached on the back of the mirror, are roughly fixed to the X-ray source in all C-arm orientations. A homography $H_{vir} \in \mathbb{R}^{3\times3}$ is estimated by the four X-ray marker point correspondances in two X-ray images such that $m_x^{ref} = H_{vir}m_x$, in which m_x is the image point in the current acquired X-ray image and m_x^{ref} is the image point in the reference X-ray image acquired in the calibrated C-arm orientation. We used the Normalized Direct Linear Transformation [8] to estimate the homography H_{vir}. Before applying the planar homography computed in the calibration of the CamC system, the X-ray image was first warped to the virtual detector plane by H_{vir} in order to guarante fixed intrinsic parameters.

4 Experiments and Results

In our experiments, the image acquisition protocol was the following, after acquiring the first X-ray image, the second X-ray image was acquired by first repositioning the C-arm with its available rotations, then performing a guided translation of the table following the feedbacks from pose estimation. Finally, the relative X-ray source motion to the table was considered as a pure rotation if the relative translation of the X-ray source was below 1 mm. All following X-ray images were acquired in the same way like the acquisition of the second image.

In order to quantify the accuracy of the panoramic X-ray images generated by our parallax-free X-ray image stitching method, we conducted an experiment to compute the difference of the overlapping area of two stitched X-ray images. One possibility to compute this difference is using similarity measurements. However, very limited feature information and high noise in the bone X-ray images make similarity measurements unreliable. This is the reason why others have employed the ruler [2] or video images [3] for X-ray image stitching. We constructed a phantom composed of spherical X-ray markers on three different planes (see figure 3(b)) and attached it to the operating table. We performed our method to stitch two X-ray images, in the overlapping area of which at least two X-ray markers on each plane were imaged. The centroids of these markers were extracted with subpixel accuracy in the two X-ray images and used to compute the difference of the overlapping area which is the distance between corresponding centroid pairs (see figure 3(b)). This procedure was performed twice. We evaluated the distances with and without applying the VDP method respectively. The overall errors were found to be 1.93 ± 1.36 pixels without VDP and 1.76 ± 1.36 pixels with VDP correction. The improvement of applying the VDP is negligible, since the X-ray source undergoes very limited rotations (around 5°).

Our parallax-free X-ray stitching was further validated in comparison to non parallax-free X-ray stitching of a cow's femoral bone. We implemented a similar method [3] for non parallax-free X-ray stitching. In the experiment, the cow bone was positioned on the operating table, and two metal tools were placed below the bone. The bone was not exactly positioned parallel to the marker pattern plane due to its big distal end. This setup is close to real clinical cases, in which it is hard to assume that a long bone is placed parallel to the marker pattern



(a) Setup (b) Panoramic X-ray Images

Fig. 4. (a) A real cow's femoral bone and the setup for generating panoramic X-ray images; (b) The top images show the parallax-free panoramic X-ray images generated by parallax-free stitching. The bottom images show the panoramic X-ray images generated by non parallax-free stitching. In each row, the left and the right panoramas were created by using the maximum pixel values and the average pixel values respectively in the overlapping areas.

plane due to the soft tissue, muscle and fat around the bone. We created the panoramic X-ray images by using maximum pixel values and average pixel values respectively in the overlapping areas. From visual inspection, the maximum yields the most uniform image but hides the misalignments (parallax effects) in the panoramas, and the average makes the individual X-ray images visible in the panoramas. Figure 4(b) shows the resulting images. A non parallax-free panoramic X-ray image was generated using the planar transformation computed for the plane (marked by red dash lines roughly in figure 4(a)) intersecting the middle of the bone horizontally and parallel to the marker pattern plane. This panoramic image has minor discontinuities in the bone boundary and clear parallax effects in the metal tools. Compared to non parallax-free stitching, parallax-free stitching produced a true parallax-free panoramic X-ray image, in which the cow bone and metal tools are almost perfectly aligned at the same time. Note that, the overlap between the consecutive images is used to illustrate the quality of the created panoramic X-ray image, and it is not necessarily required since our stitching method is independent of X-ray images.

5 Discussion and Conclusion

We presented a novel method to generate parallax-free panoramic X-ray images during surgery by using the CamC system and a planar marker pattern. However, our method is not limited to using the CamC system and can be easily applied to other C-arm setups as long as the pose of the C-arm relative to the patient's table is available, such as using an external tracking system. In our system setup, the marker pattern was attached under the operating table and is thus invisible to surgical crew during the whole operation. This allows our solution to be smoothly integrated into the surgical procedure. Our method does

not rely on overlapping X-ray regions and does not require a fronto-parallel C-arm setup. Unlike existing methods, we are able to generate a true parallax-free panoramic X-ray image that can be treated as a single image obtained from a C-arm having an exceptional wide imaging field of view. We additionally implemented a method to reduce the error caused by varying intrinsic parameters of C-arm X-ray imaging. The results show that the parallax-free panoramic X-ray image generated by our method is accurate enough and has a high visual quality. Trauma surgeons confirmed that, apart from long bone surgery, it will also support a wide range of potential intra-operative X-ray imaging applications such as pelvic fracture or whole-spine imaging, since our method is independent of the configuration of bone structures and suitable for any part of body. The table motion in our experiments is mainly translational in forward-backward direction. Currently, surgeons need put additional efforts to move the table to the correct position. However, motorized tables are already available in many clinical sites, which could further support the translation of the table. We believe that our solution for creating panoramic X-ray images has a high potential to be introduced in everyday surgical routine, and will allow surgeons to validate the quality of their treatment during orthopedic and trauma surgery with minimized radiation.

Acknowledgments. Thank to Siemens Healthcare for partial financial support.

References

1. Navab, N., Mitschke, M., Bani-Hashemi, A.: Merging visible and invisible: Two camera-augmented mobile C-arm (CAMC) applications. In: Proc. IEEE and ACM Int'l Workshop on Augmented Reality, San Francisco, CA, USA, pp. 134–141 (1999)
2. Yaniv, Z., Joskowicz, L.: Long bone panoramas from fluoroscopic x-ray images. IEEE transactions on medical imaging 23(1), 26–35 (2004)
3. Wang, L., Traub, J., Heining, S.M., Benhimane, S., Euler, E., Graumann, R., Navab, N.: Long bone x-ray image stitching using camera augmented mobile c-arm. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 578–586. Springer, Heidelberg (2008)
4. Zhang, X., Fronz, S., Navab, N.: Visual marker detection and decoding in ar systems: A comparative study. In: IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2002) (October 2002)
5. Zhang, Z.: A flexible new technique for camera calibration. Technical report, Microsoft Research (1998) (last updated in 2008)
6. Kainz, B., Grabner, M., Rüther, M.: Fast marker based c-arm pose estimation. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 652–659. Springer, Heidelberg (2008)
7. Navab, N., Mitschke, M.: Method and apparatus using a virtual detector for three-dimensional reconstruction form x-ray images (2001); Patent US 6236704; Filing date: June 30, 1999; Issue date: May 22
8. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, New York (2003)

Diffusion Tensor Field Registration in the Presence of Uncertainty

Mustafa Okan Irfanoglu[1,2], Cheng Guan Koay[2], Sinisa Pajevic[2], Raghu Machiraju[1], and Peter J. Basser[2]

[1] The Ohio State University, Columbus OH 43210, USA
[2] National Institutes of Health, NICHD, Bethesda, MD 20892, USA

Abstract. We propose a novel method for deformable tensor–to–tensor registration of Diffusion Tensor Imaging (DTI) data. Our registration method considers estimated diffusion tensors as normally distributed random variables whose covariance matrices describe uncertainties in the mean estimated tensor due to factors such as noise in diffusion weighted images (DWIs), tissue diffusion properties, and experimental design. The dissimilarity between distributions of tensors in two different voxels is computed using the Kullback-Leibler divergence to drive a deformable registration process, which is not only affected by principal diffusivities and principal directions, but also the underlying DWI properties. We in general do not assume the positive definite nature of the tensor space given the pervasive influence of noise and other factors. Results indicate that the proposed metric weights voxels more heavily whose diffusion tensors are estimated with greater certainty and exhibit anisotropic diffusion behavior thus, intrinsically favoring coherent white matter regions whose tensors are estimated with high confidence.

1 Introduction

Accurate registration of tensor fields is of great relevance in various stages of Diffusion Tensor Imaging (DTI) analysis [1]. Because of the complex nature of DTI data, cross–registration of DTI population data needed for longitudinal and multi–site studies, and brain atlas creation requires specialized mathematical tools. An accurate tensor interpolation scheme and a tensor dissimilarity metric reflecting the tensor's principal diffusivities and directions and uncertainty due to noise are needed considering the large variability among DTI data and experimental designs. Early registration approaches used derived scalar fields such as Apparent Diffusion Coefficients (ADC), Fractional Anisotropy (FA), or individual tensor components. Next–generation registration methods operated on the tensor manifolds and employed a metric to compute tensor distances such as the Riemannian [2] or Log-Euclidean [3]. Zhang *et al.* proposed a locally affine registration algorithm based on diffusion profiles, as a function of spatial direction [4]. Another work is from Cao *et al.* where the authors realize the registration by optimizing for geodesics on the space of diffeomorphisms connecting two diffusion tensor images [5]. The use of full tensor information for registration, along

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 181–189, 2009.

with metrics powerful enough to capture shape and direction information, has been shown to lead to better registration accuracy [6] [7]. However, all these approaches consider diffusion tensors as independent from the original DWIs. It is crucial to note that diffusion tensors are obtained through an optimization process on the DWIs and do not only reflect the underlying diffusion properties, but also depend on the noise in the DWIs and gradient information.

In this paper, we propose a method that uses a dissimilarity metric that not only makes use of the full estimated tensor data, but also uses the uncertainty present in typical clinical DWIs. This causes the registration to favor directionally more informative, more anisotropic and less noisy regions. To our knowledge, this property of diffusion tensors has never been investigated and employed in a registration procedure. For each voxel, a tensor–variate Gaussian distribution is constructed with a mean and a covariance matrix obtained from the tensor fitting function itself; the mean tensor provides the best estimate of the diffusion tensor while the covariance matrix quantifies the uncertainty of estimated mean diffusion tensors. The main contributions of this work are:

- using the uncertainty information present in DWIs in tensor distributions to help the registration automatically favor brain regions with high anisotropy and fiber coherence forming an internal skeleton that guides the registration.

  incorporating an initial segmentation for a tissue adaptive registration.
- providing analytically derived error differentials for faster convergence.

2 Registration Framework

The Kullback-Leibler (KL) divergence dissimilarity for tensor–variate Gaussian distributions is used as a voxelwise dissimilarity metric in a hierarchical registration framework that starts with a coarse, rigid registration, continues with affine, and finishes with a finely resolved B–splines deformable registration. A 6×6 covariance matrix is computed from the invariant Hessian of the non–linear tensor fitting function along with each mean estimated diffusion tensor to construct a Gaussian tensor–variate distribution. Figure 1 depicts the workflow of the proposed framework.

Positive Definiteness and Distributions of Diffusion Tensors. Diffusion tensors are predicted to have non–negative eigenvalues, representing the real molecular water diffusion. However, in DTI, the diffusion tensors are obtained through a physical setup not only affected by real water diffusion but also the scan parameters. This results in generally non positive-definite tensors in typical DTI scans (unconstrained fitting), especially in highly anisotropic regions such as Corpus Callossum. In his work, Pasternak et al. considers diffusion tensors as Cartesian physical quantities and shows that the Euclidean space is better suited for diffusion tensors than affine–invariant Riemannian manifolds [8]. Pajevic et. al. also shows through Monte–Carlo simulations mimicking physical imaging setups, that tensor coefficients can be modelled with a Gaussian distribution over

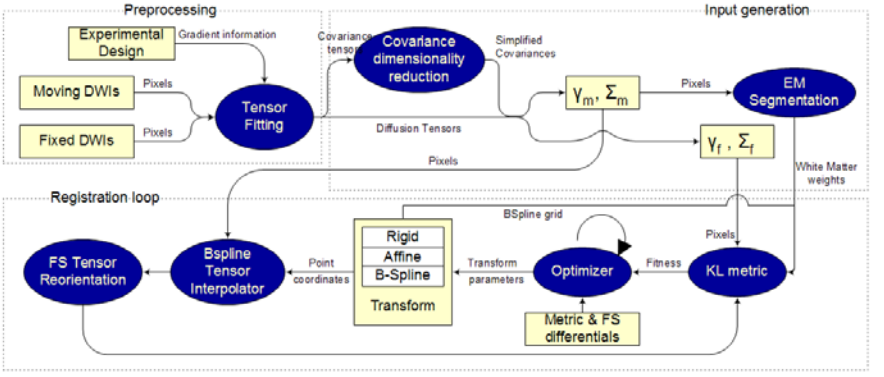


Fig. 1. The flow of the proposed registration algorithm

a wide range of SNR and the number of DWIs acquired. [9]. Aiming to cope with uncertainties such as noise and artifacts in *practical* settings, it was preferrable to use a tensor–variate Gaussian distribution in our framework, instead of a Wishart distribution, which conserve positive definiteness.

3 Methodology

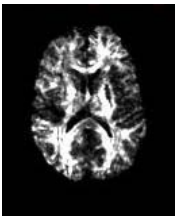
3.1 Tensor Fitting and Covariance Tensor Estimation

In a typical DTI experiment, the measured signal in a single voxel has the form [1], $s = S_0 \exp(-bg^T \mathbf{D} g)$, where the measured signal, s, depends on the diffusion encoding vector, g, the diffusion weight, b, the reference signal, S_0, and the diffusion tensor $\mathbf{D}$. Given $n \geq 7$ sampled signals derived from six non–collinear gradient directions and at least one sampled reference signal, the diffusion tensor estimate can be found with non–linear regression with the following objective function $f_{NLS}(\gamma) = \frac{1}{2}\sum_{i=1}^{n}\left(s_i - \exp\left[\sum_{i=1}^{7}\mathbf{W}_{ij}\gamma_j\right]\right)^2$. The symbol γ represents the vectorized version of diffusion tensor entries, s_i is the measured DW signal corrupted with noise, $\hat{s}_i(\gamma) = e^{\sum_{j=1}^{7}(\mathbf{W}_{ij}\gamma_j)}$ is the predicted DW signal evaluated at γ, and $\mathbf{W}$ the experimental design matrix is presented in [10].
The f_{NLS} function in Equation introduces the variability in the signal as explained in the design matrix, $\mathbf{W}$. In [11], it is shown that the diffusion tensors at each voxel can be considered as a normally distributed random variable with the covariance matrix being a function of the Hessian matrix at the optimum solution. Thus according to [11], the Hessian matrix can be computed as $\nabla^2 f_{NLS}(\gamma) = \mathbf{W}^T(\hat{\mathbf{S}}^2 - \mathbf{R}\hat{\mathbf{S}})\mathbf{W}$, where $\mathbf{S}$ and $\hat{\mathbf{S}}$ are diagonal matrices whose diagonal elements are the observed and the estimated DW signals, respectively, and $\mathbf{R} = \mathbf{S} - \hat{\mathbf{S}}$. Then, the covariance matrix of a diffusion tensor can be estimated as in [10]: $\mathbf{\Sigma}_\gamma = \sigma^2_{DW}\left[\nabla^2 f_{NLS}(\hat{\gamma})\right]^{-1}$, where σ^2_{DW} represents the variance of the noise in the DWIs [10].

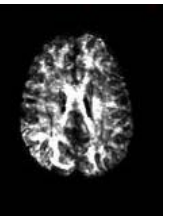
The covariance matrix is therefore a function of DWI noise , σ^2_{DW}, the gradient magnitudes and directions (embodied in the design matrix, $\mathbf{W}$) and the tissue's

$$\begin{pmatrix} \lambda+2\mu & \lambda & \lambda & 0 & 0 & 0 \\ \lambda & \lambda+2\mu & \lambda & 0 & 0 & 0 \\ \lambda & \lambda & \lambda+2\mu & 0 & 0 & 0 \\ 0 & 0 & 0 & 4\mu & 0 & 0 \\ 0 & 0 & 0 & 0 & 4\mu & 0 \\ 0 & 0 & 0 & 0 & 0 & 4\mu \end{pmatrix}$$

a)Isotropic matrix form



b)Slice 1



c)Slice 2

Fig. 2. Isotropic covariance structure in *a*) requires only 2 parameters. Deviations of the original covariance matrices from the isotropic form are displayed in b) and c). The images show that covariance matrices tend to be more anisotropic in WM regions.

underlying diffusion properties. The anisotropy, the norm, and the shape of this 6×6 matrix all provide insights on the reliability of the optimum diffusion tensor solution and the tissue properties. Figure 2 *b*) and *c*) display maps of the deviations from these matrices from the pure isotropic designs of Section 3.4, thus displaying a measure of the shape of these matrices w.r.t. the tissue type.

3.2 Dissimilarity Metric

In this work, we propose a new metric function, F, for diffusion tensor field registration. This metric uses the distribution of diffusion tensors obtained in each voxel, arising from noise and tissue properties. F is based on the symmetric KL divergence and can be described as:

$$F(I_f, I_m, \Theta) = \frac{1}{N} \sum_{p \in \Omega} w_p(I_f, I_m) \begin{pmatrix} tr(\mathbf{\Sigma}_m^{-1}\mathbf{\Sigma}_f) + (\gamma'_m - \gamma_f)^T \mathbf{\Sigma}_m^{-1}(\gamma'_m - \gamma_f) \\ + \\ tr(\mathbf{\Sigma}_f^{-1}\mathbf{\Sigma}_m) + (\gamma_f - \gamma'_m)^T \mathbf{\Sigma}_f^{-1}(\gamma_f - \gamma'_m) \end{pmatrix} \quad (1)$$

In Equation 1, γ_f signifies $\gamma_f(p)$, the diffusion tensor on the fixed image at a physical voxel location p; similarly $\mathbf{\Sigma}_f$ signifies the covariance at voxel location p, i.e., $\mathbf{\Sigma}_f(p)$, and Θ symbolizes the transformation parameters. For the moving image I_m, the covariance matrix is obtained through interpolation so $\mathbf{\Sigma}_m$ corresponds to $\mathbf{\Sigma}_m(T(p, \Theta))$. Interpolation is done through a continuous B–splines approximation framework [12]. Deforming a diffusion tensor, $\gamma_m(p)$, with a (locally) affine transformation matrix, A, involves tensor interpolation followed by reorientation. In this work, we follow the Finite Strain model proposed in [13] then the interpolated and rotated diffusion tensor $\gamma'_m(p)$ can be found to be $\gamma'_m(p) = \mathbf{R}^T \gamma_m(T(p, \Theta))\mathbf{R}$. $\mathbf{R}$ is the rotation component extracted from the affine matrix, $\mathbf{A}$, and can be found to be $\mathbf{R} = (\mathbf{A}\mathbf{A}^T)^{-1/2}\mathbf{A}$. For the elastic registration case, $\mathbf{A}$ is not constant throughout the image and can be locally estimated from the displacement field, u, as $\mathbf{A}(p) = \mathbf{I} + \mathbf{J}(u(p))$ where $\mathbf{I}$ is the identity matrix, and $\mathbf{J}(u(p))$ is the Jacobian of the deformation field at p.

Equation 1 is the Kullback-Leibler (KL) divergence symmetrized with respect to both distributions. When the first part of the equation is examined, $\frac{1}{2}(tr(\mathbf{\Sigma}_m^{-1}\mathbf{\Sigma}_f) + (\gamma'_m - \gamma_f)^T \mathbf{\Sigma}_m^{-1}(\gamma'_m - \gamma_f)$, it can be seen that the first term

in the summation, $tr(\mathbf{\Sigma}_m^{-1}\mathbf{\Sigma}_f)$, measures the similarities between the two covariance matrices; the second term is the standard Mahalanobis distance. The overall metric for the registration is the weighted $(w_p(I_f, I_m))$ summation over the KL metrics on all voxels, normalized by the number of voxels used.

3.3 Error Metric Differentials

Registration is mainly an optimization procedure, where the optimizers generally require partial differentials of the error metric with respect to the transformation parameters. Most of the DTI registration frameworks suffer from using numerical approximations to these gradients [14], such as centered differences. The problem with this approach is that it requires two metric computations per transform parameter. For deformable registrations with very large parameter space dimensionality, this approach is infeasible and an analytical solution for the differential is required. In this section, we will analytically derive the error metric gradient so that each partial differential involved has a simple form and is easy to compute numerically. This way the metric evaluations are minimized and the gradient computations are more accurate and faster. Let us have a closer look at the first term of the error metric:

$$F = \frac{1}{2}(tr(\mathbf{\Sigma}_m^{-1}(T(p,\Theta))\mathbf{\Sigma}_f(p))+(\gamma'_m(p)-\gamma_f(p))^T\mathbf{\Sigma}_m(T(p,\Theta))^{-1}(\gamma'_m(p)-\gamma_f(p)))$$

Let f be the trace term, $f = tr(\mathbf{\Sigma}_m^{-1}(T(p,\Theta))\mathbf{\Sigma}_f(p))$, and g be the Mahalanobis term, $g = (\gamma'_m(p) - \gamma_f(p))^T\mathbf{\Sigma}_m(T(p,\Theta))^{-1}(\gamma'_m(p) - \gamma_f(p))$. The differential can be expressed as $\partial F/\Theta_i = \partial f/\Theta_i + \partial g/\Theta_i$. From the chain rule, it follows that:

$$\partial f/\partial\Theta_i = \sum_{j=1}^{6}\sum_{k=1}^{6}\frac{\partial tr(\mathbf{\Sigma}_f\mathbf{\Sigma}_m^{-1}(T(p,\Theta)))}{\partial\mathbf{\Sigma}_{m\{kj\}}^{-1}}\sum_{x,y,z}\frac{\partial\mathbf{\Sigma}_m^{-1}(T(p,\Theta))}{\partial T_{x,y,z}}\frac{\partial T_{x,y,z}}{\partial\Theta_i} \qquad (2)$$

- The first differential term ,$\frac{\partial tr(\mathbf{\Sigma}_f\mathbf{\Sigma}_m^{-1}(T(p,\Theta)))}{\partial\mathbf{\Sigma}_{f\{kj\}}^{-1}}$, is just $\mathbf{\Sigma}_{f\{kj\}}$ from the symmetry of covariance matrices and the derivative of traces w.r.t the matrices. Also note that the inverses of the covariance matrices are stored and used as images, cancelling the need for the inverse operation for the differential. Additionally, as explained in Section 3.4 the isotropic covariance matrix $\mathbf{\Sigma}_m^{-1}(T(p,\Theta))$ is obtained only using interpolation but not reorientation due to rotational invariance assumption, yielding a simpler formula [9].

  The second partial in Equation 2 represents the image gradient of the maps of each covariance components w.r.t. imaging directions. These gradients need to be computed once at the beginning of the registration.
- The last term $\frac{\partial T_{x,y,z}}{\partial\Theta_i}$ corresponds to the Jacobian of the transformation and needs to be computed once per iteration.

The Mahalanobis part of the function F, i.e., the function g, has a more complicated differential due to the rotation of diffusion tensors $\gamma_m(T(p,\Theta))$ into $\gamma'(p)$ if an affine or deformable registration scheme is employed. Let a be $a = \gamma'_m - \gamma_f$,

then the Mahalanobis part g can be rewritten as $g = \sum_j \sum_k a_j a_k \mathbf{\Sigma}^{-1}_{m\{jk\}}$. Then the differential can be rewritten as:

$$\frac{\partial g}{\partial \Theta_i} = \sum_j \sum_k \frac{\partial a_j}{\partial \Theta_i} a_k \mathbf{\Sigma}^{-1}_{m\{jk\}} + \sum_j \sum_k a_j \frac{\partial a_k}{\partial \Theta_i} \mathbf{\Sigma}^{-1}_{m\{jk\}} + \sum_j \sum_k a_j a_k \frac{\partial \mathbf{\Sigma}^{-1}_{m\{jk\}}}{\partial \Theta_i}$$

The differential in the last term, $\frac{\partial \mathbf{\Sigma}^{-1}_{m\{jk\}}}{\partial \Theta_i}$ is the same as the one used in Equation 2, i.e., $\sum_{x,y,z} \frac{\partial \mathbf{\Sigma}^{-1}_m (T(p,\Theta))}{\partial T_{x,y,z}} \frac{\partial T_{x,y,z}}{\partial \Theta_i}$. With the finite strain model, a can be described as $a = R^T \gamma_m(T(p,\Theta))R - \gamma_f$. Then the first differential becomes:

$$\frac{\partial a_i}{\partial \Theta_z} = \frac{\partial R^T}{\partial \Theta_z} \gamma_m(T(p,\Theta))R + R^T \frac{\partial \gamma_m(T(p,\Theta))}{\partial \Theta_z} R + R\gamma_m(T(p,\Theta)) \frac{\partial R}{\partial \Theta_z} \quad (3)$$

The second partial in Equation 3 can be found similarly to the covariance matrix case and is $\sum_{x,y,z} \frac{\partial \gamma_m(T(p,\Theta))}{\partial T_{x,y,z}} \frac{\partial T_{x,y,z}}{\partial \Theta_i}$. In the case of an affine transformation, where the parameters Θ_i corresponds to the entries in the affine matrix, A, the partial derivative of the rotation matrix, R, with respect to the transformation parameter, Θ_z, comes from the chain rule, $\frac{\partial R}{\partial \Theta} = \frac{\partial R}{\partial A}$. For the elastic registration case, the local affine matrix is estimated from the displacement field, u, and the differential becomes $\frac{\partial R}{\partial \Theta_z} = \sum_j \sum_k \frac{\partial R}{\partial u_{jk}} \frac{\partial u_{jk}}{\partial \Theta_z}$. For B–splines registration of order 3, the displacement field can be written as, $u(p,\beta) = \sum_i \sum_j \sum_k \beta_{ijk} b_{i,3}(p_x) b_{j,3}(p_y) b_{k,3}(p_z)$, where β_{ijk} are B–splines weights corresponding to parameters Θ and $b_{.,3}$ are 3^{rd} order spline basis functions. Then the second partial derivative, $\frac{\partial u_{jk}}{\partial \Theta_z}$, is just $b_{i,3}(p_x) b_{j,3}(p_y) b_{k,3}(p_z)$ for $\Theta_z = \beta_{ijk}$. The first term, $\frac{\partial R}{\partial u_{jk}}$, can be found in in [14].

3.4 Covariance Matrix Dimensionality Reduction

Independent components of diffusion tensors (6) and covariance matrices (21) generally yield a total of 27 dimensions, which poses problems in terms of memory and speed during registration. Being a function of the matrix W, the covariance matrix's form depends on the number of gradients and the direction of gradients used. In [9], it is shown that with sufficient number of diffusion gradients sampling the unit icosahedron densely enough, the $4D$ covariance tensor $(3 \times 3 \times 3 \times 3)$ corresponding to the $2D$ covariance matrix tends to be isotropic and rotationally invariant. These isotropic covariance matrix yield a specific 6×6 matrix structure, with the block matrix form shown in Figure 2 a).

3.5 Tissue Segmentation

The tensor covariance matrix provides additional information on the tissue type. To further use this additional information for a more robust and faster registration, we perform a classical Expectation–Maximization (EM) segmentation initialized with K-means clustering, with the distance function originating from

our KL-metric and tensor–variate distributions derived from the mean and covariance matrices. This procedure is used in the registration initialization. For each moving image, first a segmentation is carried out. The probability of a voxel being a WM voxel obtained from the EM segmentation is used as the weighting factor $w_p(I_f, I_m)$ in Equation 1. Additionally, the segmentation labels are used to build a multi–level grid for B–splines registration. A coarser B–splines transformation grid is placed on CSF locations to decrease the computational complexity, whereas a denser grid is used for WM.

4 Experiments and Results

We acquired data from six healthy subjects with DTI parameters, $b = 1000 s/mm^2$, 72 diffusion gradient directions. Matrix sizes for all images were 128×157 with 114 axial slices and $1.5mm$ isotropic voxel resolution. One of the images was chosen to be the fixed image and the other five were used as moving images. For comparison, we implemented a benchmark multi–channel registration algorithm with *six* channels for tensor components, including one channel for FA and one channel for ADC. The benchmark method followed the same vector image registration steps. Standard deviation maps of the FA maps were also computed from the registered images. For the described dataset, the proposed registration pipeline with rigid, affine, and B–splines transformations (maximum grid size $20 \times 20 \times 20$), takes on the average 30 minutes per image on a modern computer.

4.1 Segmentation Outputs

We segmented a brain image by using three different levels of information: the isotropic covariance matrices (trace part of the error metric), only the full covariance matrices, and full covariance matrices along with the diffusion tensors as described by the error metric. Figure 3 displays the result of these segmentations. The images in Figure 3 show that the tensor covariance information brings additional information about the tissue type. With increasing complexity of the covariance matrix structure, tissue layers can better be discriminated . The use of full covariance matrix along with the diffusion tensor further improves the segmentation (Figure 3 (d)).

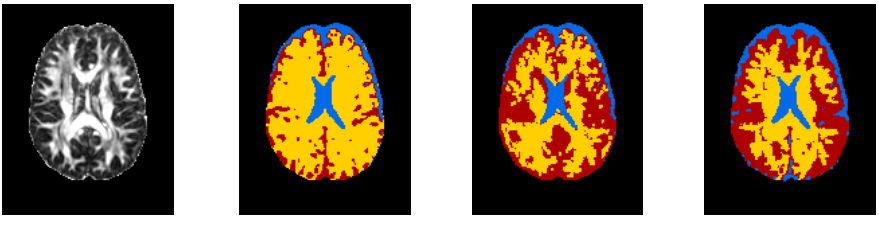


Fig. 3. Segmentation results. (a) Fa image. (b)Segmentation with only isotropic covariance matrix. (c) Segmentation using full covariance matrix. Segmentation of white matter improves using full covariance matrix and diffusion tensor information (d).

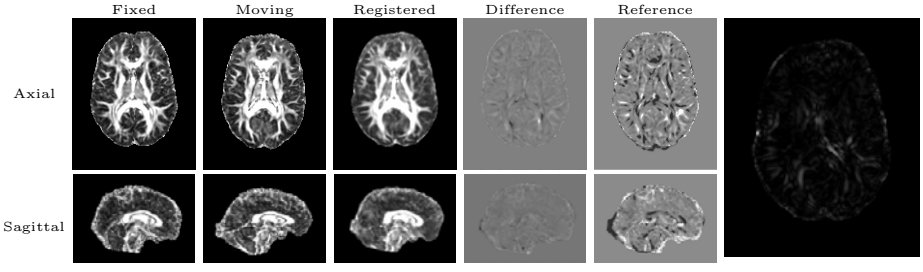


Fig. 4. Output of the registration algorithm. The fixed, moving, registered moving, difference images, and FA standard deviation displayed in different columns.

4.2 Registration Results

Figure 4 displays the output of the registration algorithm. The metric proves to perform well on white matter regions, as can be observed from the similarity of the images in the first and third columns. The difference image of the registered moving image and the fixed image is displayed on the fourth column, where it is visible that the metric performs significantly better than the benchmark method on the Corpus Callosum. The difference image for the benchmark method is displayed in the fifth column. The sixth column, displays the standard deviation of the FA maps of the five images registered with the proposed method. Note that the performance of the algorithm on white matter is clearly visible.

5 Discussions and Conclusions

In this work, we proposed a novel, robust and fast approach for tensor–to–tensor registration for Diffusion Tensor Image data, suitable for group analysis and tensor atlasing problems. The proposed metric captures the uncertainty of the diffusion tensors with a tensor-variate Gaussian distribution. Our future research directions include the analysis of the shape and isotropy characteristics of covariance matrices and testing of the algorithm with a larger population.

References

1. Basser, P.J., Jones, D.K.: Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. NMR Biomed. 15(7-8), 456–467 (2002)
2. Fletcher, P., Joshi, S.: Principal geodesic analysis on symmetric spaces: Statistics of diffusion tensors. In: Sonka, M., Kakadiaris, I.A., Kybic, J. (eds.) CVAMIA/MMBIA 2004. LNCS, vol. 3117, pp. 87–98. Springer, Heidelberg (2004)
3. Arsigny, V., Fillard, P., Pennec, X.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine 56, 411–421 (2006)
4. Zhang, H., Yushkevich, P.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 172–179. Springer, Heidelberg (2005)

5. Cao, Y., Miller, M., Mori, S., Winslow, R.L., Younes, L.: Diffeomorphic matching of diffusion tensor images. In: CVPR Workshop, vol. 22(17), pp. 67–67 (2006)
6. Jinzhong, Y., Dinggang, S., Ragini, V.: Diffusion tensor image registration using tensor geometry and orientation features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 905–913. Springer, Heidelberg (2008)
7. Zhang, H., Avants, B.B., Yushkevich, P.A., Woo, J.H., Wang, S., McCluskey, L.F., Elman, L.B., Melhem, E.R., Gee, J.C.: High-dimensional spatial normalization of diffusion tensor images improves the detection of white matter differences: An example study using amyotrophic lateral sclerosis. IEEE Trans. Med. Imaging 26(11), 1585–1597 (2007)
8. Pasternak, O., Verma, R., Sochen, N., Basser, P.J.: On what manifold do diffusion tensors live. In: MICCAI Workshop (2008)
9. Basser, P.J., Pajevic, S.: A normal distribution for tensor-valued random variables: applications to diffusion tensor mri. IEEE TMI 202, 785–794 (2003)
10. Koay, C.G., Chang, L., Pierpaoli, C., Basser, P.J.: Error propagation framework for diffusion tensor imaging via diffusion tensor representations. IEEE Transactions on Medical Imaging 26(8), 1017–1034 (2007)
11. Koay, C.G., Chang, L., Pierpaoli, C., Basser, P.J.: A unified theoretical and algorithmic framework for least square methods of estimation in diffusion tensor imaging. Journal of Magnetic Resonance 182, 115–125 (2006)
12. Pajevic, S., Aldroubi, A., Basser, P.J.: Continuous tensor field approximation of diffusion tensor mri data. Visualization and Processing of Tensor Fields (2006)
13. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. TMI 20(11), 1131–1139 (2001)
14. Yeo, B., Vercauteren, T., Fillard, P., Pennec, X., Ayache, N., Clatz, O.: DTI registration with exact finite-strain differential. In: IEEE ISBI (2008)

Non-rigid Registration of High Angular Resolution Diffusion Images Represented by Gaussian Mixture Fields

Guang Cheng[1], Baba C. Vemuri[1,⋆], Paul R. Carney[2], and Thomas H. Mareci[3]

[1] CISE, University of Florida
{gcheng,vemuri}@cise.ufl.edu
[2] Department of Pediatrics, University of Florida
carnepr@peds.ufl.edu
[3] Department of Biochemistry and Molecular Biology, University of Florida
thmareci@ufl.edu

Abstract. In this paper, we present a novel algorithm for non-rigidly registering two high angular resolution diffusion weighted MRIs (HARDI), each represented by a Gaussian mixture field (GMF). We model the non-rigid warp by a thin-plate spline and formulate the registration problem as the minimization of the L2 distance between the two given GMFs. The key mathematical contributions of this work are, (i) a closed form expression for the derivatives of this objective function with respect to the parameters of the registration and (ii) a novel and simpler re-orientation scheme based on an extension to the "Preservation of Principle Directions" technique. We present results of our algorithm's performance on several synthetic and real HARDI data sets.

1 Introduction

Diffusion-Weighted MR Imaging (DWMRI) is a unique non-invasive technique that makes the MR signal sensitive to the water molecule diffusion in the imaged tissue and infer its structure *in vivo*. When using DWMRI in longitudinal studies, multi-subject studies, and such tasks, non-rigid registration is essential to quantify the differences between the acquired data sets.

The diffusivity function at a lattice point in a DWMRI when approximated by a rank-2 symmetric positive definite tensor leads to what is popularly dubbed a diffusion tensor image (DTI)[1]. Several DTI registration methods have been reported in literature to date. Alexander et al. [2] developed a registration technique for DTI and observed that a re-orientation of the tensors was necessary as part of the registration of the DWMRI data sets. Some existing methods [3,4,5] used one or more rotation invariant characteristics of diffusion tensors to perform the registration and thereby avoiding the re-orientation task at each iteration of the registration algorithm. A variational model for the diffeomorphic DTI registration was proposed in [6] along with an analytic gradient of the Preservation

⋆ Corresponding author.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 190–197, 2009.

of Principle Direction (PPD) reorientation strategy, which is very complex and cumbersome requiring substantial compute time. In [7] the analytic gradient of the Finite Strain (FS) strategy was developed for diffeomorphic DTI registration. None of these methods are directly applicable to the case of registering GMFs.

Rank-2 SPD tensors are a good approximation for the single fiber geometry. However, this model is known to fail at the locations with complex tissue geometry such as fiber crossings. Some higher order models [8,9,10,11,12] have been reported in literature to tackle this problem. Research on registration of the resulting representations from these methods has not been addressed to date with the exception of work in [13] wherein, a piece-wise affine approximation to the nonrigid registration of fourth order tensor fields was developed. Most recently, a fluid-flow based model of the non-rigid transformation was introduced in [14] for registering two given DTI data sets. Authors claimed that their work was easily extendable to registration of higher order representations derived from HARDI data but no such extension has been reported to date.

In this paper, we present a novel non-rigid registration method for HARDI datasets represented by GMFs generated by using the algorithm described in [9]. The non-rigid registration between the GMFs in this paper is represented by a thin plate spline (TPS) and the dissimilarity between the GMFs being registered is expressed by the $L2$ distance between Gaussian mixtures at corresponding points between the image lattices, and can be viewed as an extension of the work in [15]. The re-orientation operation is handled by a novel and significantly improved extension (developed here) of the PPD strategy in [2]. The key contributions of the work reported here are, (i) it is the first attempt at the non-rigid registration of HARDI data sets represented by GMFs. (ii) An objective function involving the L2 distance between Gaussian mixtures is used that leads to a closed form expression for the distance. (iii) A significant extension of the PPD strategy is to handle re-orientation of Gaussian mixtures involving a derivation of the objective function and its analytic gradient both of which are in a much simpler form than those derived for DTI re-orientation in [6]. Experiments on synthetic and real data along with comparisons demonstrate the performance of our algorithm.

The rest of the paper is organized as follows: In section 2, we present registration framework including a derivation of the analytic form of the L2 distance between two Gaussian mixtures and our re-orientation strategy. We then present the TPS formulas followed by the derivation of a closed form expression of the cost function derivatives w.r.t. the deformation parameters. In section 3, we present synthetic and real data experiments along with comparisons to other methods and conclude in section 4.

2 Non-rigid Registration of Gaussian Mixture Fields

There are several ways to represent the HARDI data and the choice depends on the end goal. We choose to employ the diffusion propagator *pdf* $P(\mathbf{r}, t)$ – where $\mathbf{r}$

is the displacement vector from a given location at time t – as our representation and compute this field from the given HARDI data. The reason for this choice is the generality and accuracy of the continuous mixture model used here & described in [8,9].

We now present our registration framework for GMFs obtained using the method presented in [9]. Here a GMF is a map from R^3 to the space of smooth functions C^∞. Given the target GMF I and the source GMF J, registration can be described as a process of estimating the transform $T : R^3 \rightarrow R^3$ from the coordinates of I to the coordinates of J, which best aligns the two GMFs. This can be done by minimizing the dissimilarity between the target GMF and the transformed source GMF with respect to T. The energy function for this optimization process can be written as

$$E(T) = \int_{R^3} dist^2(I_T(\mathbf{x}), J(T(\mathbf{x})))d\mathbf{x} \tag{1}$$

where $I_T(\mathbf{x})$ and $J(T(\mathbf{x}))$ for a fixed position $\mathbf{x} \in R^3$ are two zero mean Gaussian mixtures(GMs), $dist(\circ, \circ)$ denotes the dissimilarity measure between the two GMs and I_T denotes the GMF I after re-orientation. The details of equation 1 are described in the remaining of this section.

L_2 Distance Between Zero Mean Gaussian Mixtures: In this paper we use the L_2 distance as a measure of the dissimilarity between zero mean GMs. Let $f(\boldsymbol{r}) = \Sigma_{i=1}^{M}\eta_i G(\boldsymbol{r}; 0, \boldsymbol{\Sigma}_i)$ and $g(\boldsymbol{r}) = \Sigma_{j=1}^{N}\rho_j G(\boldsymbol{r}; 0, \boldsymbol{\Gamma}_j)$ be two GM density functions, where $\mathbf{r} \in R^3$ is the displacement vector and η_i, ρ_j denote the mixture weights of the corresponding Gaussian components $G(\boldsymbol{r}; 0, \boldsymbol{\Sigma}_i)$ and $G(\boldsymbol{r}; 0, \boldsymbol{\Gamma}_j)$ with covariance matrices $\boldsymbol{\Sigma}_i$ and $\boldsymbol{\Gamma}_j$ respectively. The L_2 distance between f and g can be written as a quadratic function of the mixture weights

$$dist^2(f, g) = \int_{R^3} (f(\mathbf{r}) - g(\mathbf{r}))^2 d\mathbf{r} = \boldsymbol{\eta}^t \boldsymbol{A}\boldsymbol{\eta} + \boldsymbol{\rho}^t \boldsymbol{B}\boldsymbol{\rho} - 2\boldsymbol{\eta}^t \boldsymbol{C}\boldsymbol{\rho} \tag{2}$$

where $\boldsymbol{\eta} = (\eta_1, ..., \eta_M)^t$ and $\boldsymbol{\rho} = (\rho_1, ..., \rho_N)^t$ are vectors representing the mixture weights, and $\boldsymbol{A}_{M\times M}, \boldsymbol{B}_{N\times N}$ and $\boldsymbol{C}_{M\times N}$ are the matrices generated by the Gaussian components, with their elements to be $A_{i_1 i_2} = ((2\pi)^3 det(\boldsymbol{\Sigma}_{i_1} + \boldsymbol{\Sigma}_{i_2}))^{-1/2}$, $B_{j_1 j_2} = ((2\pi)^3 det(\boldsymbol{\Gamma}_{j_1} + \boldsymbol{\Gamma}_{j_2}))^{-1/2}$ and $C_{ij} = ((2\pi)^3 det(\boldsymbol{\Sigma}_i + \boldsymbol{\Gamma}_j))^{-1/2}$

If a 3D covariance matrix $\boldsymbol{\Sigma}$ has eigen values $\lambda_1 \geq \lambda_2 = \lambda_3$ (cylindrical symmetry), its rank one decomposition can be written as $\boldsymbol{\Sigma} = (\lambda_1 - \lambda_2)\mathbf{u}\mathbf{u}^t + \lambda_2\mathbf{I}$, where $\mathbf{u}$ is the principal eigen vector of $\boldsymbol{\Sigma}$. Given another cylindrically symmetric covariance matrix $\boldsymbol{\Gamma} = (\xi_1 - \xi_2)\mathbf{v}\mathbf{v}^t + \xi_2\mathbf{I}$, we can derive the following equation after applying some matrix algebra,

$$det(\boldsymbol{\Sigma} + \boldsymbol{\Gamma}) = \alpha - \beta(\mathbf{u}^t\mathbf{v})^2 \tag{3}$$

where $\alpha = (\xi_1 + \lambda_2)(\lambda_1 + \xi_2)(\lambda_2 + \xi_2), \beta = (\lambda_1 - \lambda_2)(\xi_1 - \xi_2)(\lambda_2 + \xi_2)$. It is reasonable to assume that all the Gaussian components in our GMFs are cylindrically symmetric, as the fibers have approximately cylindrical geometry. Thus, equation 3 can be used to compute the matrices $\mathbf{A}, \mathbf{B}$ and $\mathbf{C}$ in equation 2. Using equation 3, reorientation can be easily achieved and the registration function along with its gradient can be simplified.

Re-orientation: As a GMF is not rotation invariant, we need to perform a re-orientation during the registration process. In this context, DTI reorientation methods would be useful to consider and generalize. There are three DTI re-orientation strategies reported in literature, namely, (i) Preservation of Principle Directions(PPD) [2], (ii) Finite Strain(FS) [2] and (iii) Re-transformation [13].

In the PPD, the principle direction $\mathbf{v}$ of the 2rd order tensor $\mathbf{D}$ is transformed to $\hat{\mathbf{v}} = Fv/|Fv|$ where $\mathbf{F}$ is the Jacobian of the transform T. In this paper, we extend the PPD strategy to the mixture of Gaussians model in a direct way by applying it to the covariance matrices of each Gaussian component. With the cylindrical geometry assumption on the nerve fibers, let the source GMF $J(\mathbf{x}) = \Sigma_{i=1}^{M}\eta_i(\mathbf{x})G(\boldsymbol{r}; 0, (\lambda_1^i - \lambda_2^i)\boldsymbol{u}_i\boldsymbol{u}_i^t + \lambda_2^i\mathbf{I})$ and target GMF $I(\mathbf{x}) = \Sigma_{j=1}^{N}\rho_j(\mathbf{x})G(\boldsymbol{r}; 0, (\xi_1^j - \xi_2^j)\boldsymbol{v}_j\boldsymbol{v}_j^t + \xi_2^j\mathbf{I})$. Here we can assume $\lambda^i, \xi^j, \mathbf{u_i}, \mathbf{v_j}$ to be constant across the lattice for simplicity. The re-orientation is applied to the target GMF I by applying the Jacobian $\mathbf{F}$ to each of its covariance matrices. Thus, $I_T(\mathbf{x}) = \Sigma_{j=1}^{N_x}\rho_j(\mathbf{x})G(\boldsymbol{r}; 0, (\xi_1^j - \xi_2^j)\boldsymbol{v_j}\boldsymbol{F}\boldsymbol{F^t}\boldsymbol{v}_j^t/||\mathbf{F}\mathbf{v}_i||^2 + \xi_2^j\mathbf{I})$. Using Equation 3, the energy function can be written as

$$E(T) = \int_{R^3} \boldsymbol{\eta}(\mathbf{T}(\mathbf{x}))^t\boldsymbol{A}\boldsymbol{\eta}(\mathbf{T}(\mathbf{x})) + \boldsymbol{\rho}(\mathbf{x})^t\boldsymbol{B_T}\boldsymbol{\rho}(\mathbf{x}) - 2\boldsymbol{\eta}^t(\mathbf{T}(\mathbf{x}))\boldsymbol{C_T}\boldsymbol{\rho}(\mathbf{x})d\mathbf{x} \quad (4)$$

where $C_{Tij} = k(\iota_{ij})$ and $B_{Tj_1j_2} = k(\tau_{j_ij_2})$, with scalar $\iota_{ij} = (\mathbf{u}_i^t\mathbf{F}\mathbf{v}_j)^2/||\mathbf{F}\mathbf{v}_j||^2$, $\tau_{j_ij_2} = (\mathbf{v}_{j_1}^t\mathbf{F^t}\mathbf{F}\mathbf{v}_{j_2})^2/(||\mathbf{F}\mathbf{v}_{j_1}||^2||\mathbf{F}\mathbf{v}_{j_2}||^2)$, and $k(t) = ((2\pi)^3(\alpha - \beta t))^{(-1/2)}$, α and β are scalar constants defined using the eigen values as in equation 3.

2.1 The TPS-Based Non-rigid Registration

We are now ready to present the estimation of the non-rigid registration T – between the GMFs I ad J – represented by a TPS [15]. Let the set of control grid points for computing the registration be $\{\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_n\}$, the transform is given by$T(\mathbf{x}) = \mathbf{WH}(\mathbf{x}) + \mathbf{Ax} + \mathbf{t}$, where $\mathbf{Ax} + \mathbf{t}$ is the affine part, and $\mathbf{WH}(\mathbf{x})$ is the non-rigid part. $\mathbf{H}(\mathbf{x}) = (H_1(\mathbf{x}), H_2(\mathbf{x})..,H_n(\mathbf{x}))^t$ is a group of non-linear kernel functions with $H_i(\mathbf{x}) = r_i(\mathbf{x})^2\log(r_i(\mathbf{x}))$ in the case of 2D and $H_i(\mathbf{x}) = r_i(\mathbf{x})^3$ in the case of 3D, where $r_i(\mathbf{x}) = ||\mathbf{x}-\mathbf{x}_i||$. This non-rigid transform is regularized by minimizing the bending energy of the TPS given by $trace(\mathbf{WKW}^t)$. Here $\mathbf{K}$ is the dissimilarity matrix with elements $K_{ij} = r_i(\mathbf{x}_j)$. To perform re-orientation, the Jacobian matrix $\mathbf{F}$ at every image grid point also needs to be computed, and is given by $\mathbf{F} = \mathbf{WJ}(\mathbf{H}(\mathbf{x})) + \mathbf{A}$, where $\mathbf{J}(\mathbf{H}(\mathbf{x})) = (\nabla {H_1}^t, \nabla {H_2}^t, ...\nabla {H_n}^t)^t|_x$ is the Jacobian of $\mathbf{H}(\mathbf{x})$. The non-rigid registration here involves minimization of the energy function w.r.t. the parameters $\mathbf{A}, \mathbf{t}$ and $\mathbf{W}$. The affine parameters $\mathbf{A}$ and $\mathbf{t}$ can be solved for via an affine registration technique prior to solving for the non-rigid component $\mathbf{W}$.

2.2 Analytic Derivative of the Objective Function

The derivative of the objective function is straight forward to compute by applying the chain rule. The partial derivative of the energy function (without

the bending energy) w.r.t. the elements w_{kc} in the matrix $\mathbf{W}$ (recall that $\mathbf{W}$ is weight matrix of the non-rigid part of the TPS) can be written as:

$$\frac{\partial E}{\partial w_{kc}} = \int_{R^3} 2\frac{\partial \mathbf{T}^t}{\partial w_{kc}}\nabla\boldsymbol{\eta}^t\boldsymbol{A}\boldsymbol{\eta} - 2\frac{\partial \mathbf{T}^t}{\partial w_{kc}}\nabla\boldsymbol{\eta}^t\boldsymbol{C_T}\boldsymbol{\rho} + \boldsymbol{\rho}^t\frac{\partial \boldsymbol{B_T}}{\partial w_{kc}}\boldsymbol{\rho} - 2\boldsymbol{\eta}^t\frac{\partial \boldsymbol{C_T}}{\partial w_{kc}}\boldsymbol{\rho}d\mathbf{x}$$

where the derivative of the TPS function $\frac{\partial \mathbf{T}}{\partial w_{kc}}$ is a 3-by-1 vector, with its k-th element being the c-th kernel function value at x, namely $H_c(x)$, and two other elements being zeros; the derivative of $\mathbf{C_T}$ and $\mathbf{B_T}$ can be computed as:

$$\boldsymbol{\rho}^t\frac{\partial \boldsymbol{B_T}}{\partial w_{kc}}\boldsymbol{\rho} = Tr(\frac{\partial \mathbf{F}^t}{\partial w_{kc}}\sum_{j_1,j_2}\rho_{j_1}\rho_{j_2}\frac{\partial B_{j_1j_2}}{\partial \mathbf{F}}); \quad \boldsymbol{\eta}^t\frac{\partial \boldsymbol{C_T}}{\partial w_{kc}}\boldsymbol{\rho} = Tr(\frac{\partial \mathbf{F}^t}{\partial w_{kc}}\sum_{i,j}\eta_i\rho_j\frac{\partial C_{ij}}{\partial \mathbf{F}})$$

where $\frac{\partial B_{j_1j_2}}{\partial \mathbf{F}} = 2k'\tau_{j_1j_2}(\mathbf{F}(\mathbf{v}_{j_i}\mathbf{v}_{j_2}^t + \mathbf{v}_{j_2}\mathbf{v}_{j_1}^t)/(\mathbf{v}_{j_1}^t\mathbf{F}^\mathbf{t}\mathbf{F}\mathbf{v}_{j_2}) - \mathbf{F}\mathbf{v}_{j_1}\mathbf{v}_{j_1}^t/||\mathbf{F}\mathbf{v}_{j_1}||^2 - \mathbf{F}\mathbf{v}_{j_2}\mathbf{v}_{j_2}^t/||\mathbf{F}\mathbf{v}_{j_2}||^2)$, $\frac{\partial C_{ij}}{\partial \mathbf{F}} = 2k'\iota_{ij}(\mathbf{u}_i\mathbf{v_j}^t/\mathbf{u}_i^t\mathbf{F}\mathbf{v}_j - \mathbf{F}\mathbf{v}_j\mathbf{v_j}^t/||\mathbf{F}\mathbf{v}_j||^2)$, and $\frac{\partial F}{\partial w_{kc}} = \boldsymbol{y}^{kc}$, where $\boldsymbol{y}^{kc}$ is a 3-by-3 matrix with each element given by $y_{nm}^{kc} = \frac{\partial H_c}{\partial x_m}\delta(n-k)$ and $\delta\cdot$ being the discrete delta function.

Since we have the gradient of the energy function in the analytic form, any of the gradient based optimization strategies in the literature can be applied to optimize the registration cost function. In this paper we use the efficient limited-memory Broyden-Fletcher-Goldfarb quasi-Newton technique [16]. Note that, as claimed in [9], the weight vectors in the Gaussian mixtures are sparse. Thus, we can just compute the elements of the matrices corresponding to nonzero weights.

3 Experiments

We now report the experimental results obtained on synthetic and real HARDI datasets. In the synthetic data experiment, we registered two images with crossing fiber bundles. Firstly, a 2D synthetic image with two crossing fiber bundles was manually generated. Then, 20 randomly deformed images were generated from the crossing bundles by applying a b-spline based non-rigid deformation and a PPD based re-orientation scheme described here. The method described in [17] was used to generate the simulated MR signals from the fiber bundles. Rician noise was added to simulate data at 4 different noise levels with $SNR = 50, 20, 10$ and 5. The method in [9] was used to generate the GMF from the MR signals with 46 Gaussian components at each voxel. After the data generation, we registered each of the randomly deformed source images to the fixed target image separately. To evaluate the registration, the resulting deformation obtained from the registration was applied to the noise free source image, and then the relative dissimilarity between the deformed source and target image was computed as the relative error in registration given by,

$$Err = \frac{\sum_{\mathbf{x}\in\mathbf{R}}\sqrt{\int_{S^2}(I_\mathbf{x}(\mathbf{r}) - J'_\mathbf{x}(\mathbf{r}))^2 d\mathbf{r}}}{\sum_{\mathbf{x}\in\mathbf{R}}\sqrt{\int_{S^2}(I_\mathbf{x}(\mathbf{r}))^2 d\mathbf{r}}} \tag{5}$$

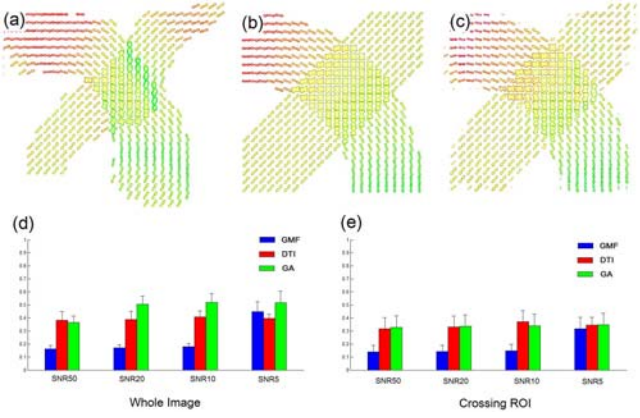


Fig. 1. Experimental results on synthetic dataset. (a) is the source, (b) is the target and (c) is the deformed source image. Figure (d) and (e) are the mean and standard deviation of the error for the 20 registrations from all the three methods at different noise levels computed using equation 5 in the two different ROIs.

where $\mathbf{R}$ is a user selected region to evaluate the errors; I, J' denotes the target and deformed source image; $I_{\mathbf{x}}(\mathbf{r})$ and $J'_{\mathbf{x}}(\mathbf{r})$ are the displacement probability profiles (represented in spherical harmonic coefficients) at voxel $\mathbf{x}$. Note that this dissimilarity measure is different from the one used in our registration algorithm. In this paper, two different regions are used for error computation: (1) The whole fiber bundle region and (2) The region containing just the fiber crossings. Also, a GA (generalized Anisotropy) based registration algorithm (using a sum of squared differences cost) and the DTI based registration algorithm in [18] were tested using the same data set for comparison purposes. The data sets and the results are displayed in Figure 1. Figure (f) and (g) show that our method has much lower mean and standard deviation of registration errors for both the chosen ROIs for first three noise levels, and has a comparable error to other two methods at $SNR5$ (which is very high amount of noise).

In the real data experiment, we used rat brain DWMRI. For data acquisition, in each scan 27 diffusion gradients were used (with Δ and δ set to be $17.6ms$ and $4.8ms$) with a $b = 800s/mm^2$ for 21 of the 27 gradients and the remaining with a $b = 100s/mm^2$. The image size was $100 \times 100 \times 12$, with each voxel of size, $0.3mm \times 0.3mm \times 0.9mm$. In the first experiment, we show results from two distinct rats. For the second experiment, we depict results from a single rat, taken at different time points after implanting electrodes used to stimulate injury that results in the development of epilepsy. Consecutive scans were used in this registration experiment to capture structural changes via the non-rigid registration.

For the first real data experiment, the data and the registration result are shown in a checker-board view in Figure 2. (In the checker-board view alternate blocks come from the source and target images respectively). From the figure its evident that our registration algorithm successfully aligned the two given HARDI data sets. In sub-figure (e) and (f), there is a fiber bundle from the

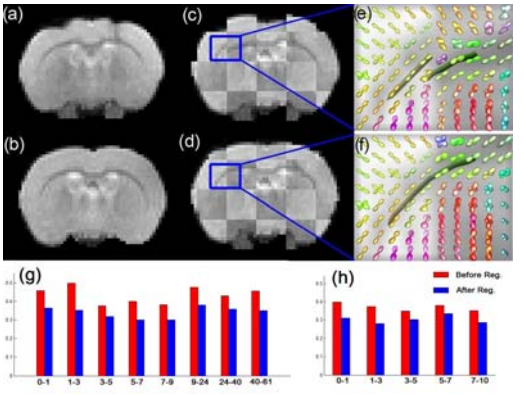


Fig. 2. Real data experiments:(a) and (b) are the source and targets S_0 images, (c) checker-board view of source and target images, (d) checker-board view of the deformed source and target images. (e) and (f) Displacement probabilities in the ROI shown in (c) and (d) respec. Misaligned fiber tracts depicted in (e) are well aligned after the registration shown in (f). (g) and (h) are error bars of consecutive HARDI scans with red and blue bars denoting the error before and after registration. The labels on the X-axis denote the scan dates of the input image pairs.

corpus callosum in the ROI, which contains most of the probability profile in green and brown (the stream tubes are a result of the tracking results using a simple vector field integration based tracker). In (e) (before registration) this fiber bundle is shown disrupted and depicted as two distinct disconnected tracts in the checker-board view, and in (f) (after registration) it is shown as one coherent connected tract indicating the accuracy of the registration visually.

For quantitative evaluation, we compared the dissimilarity of the target and source image pair before and after registration using equation 5. Then, we applied our algorithm to consecutive DWMR time scans of the same rat, and plotted the dissimilarity for the image pairs before and after registration. Plates (g) and (h) in Figure 2 are the results for two different rats one with 9 consecutive scans and the other with 6. From the figure, it is apparent that the dissimilarity between them has decreased significantly after registration. This indicates the effectiveness of our registration method on real data.

4 Conclusion

In this paper we presented a novel non-rigid registration algorithm for HARDI data represented by GMFs. The non-rigid transformation is modeled using a TPS function. The key contributions of this work are, (i) it is the first report on non-rigid registration of GMFs representing HARDI data sets. (ii) A non-trivial derivation leading to a closed form expression for the gradient of the registration was presented. (iii) A novel extension of the PPD strategy of re-orientation and a novel derivation of the reorientation that is simpler than the previously reported closed form for PPD-based re-orientation of DTI. Both real and synthetic data

experiments along with comparisons were presented depicting the effectiveness of the presented registration algorithm.

Acknowledgment. This research was funded by the NIH grants EB007082 to BCV, & EB004752 to PC & TM.

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson. B 103, 247–254 (1994)
2. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. IEEE TMI 20, 1131–1139 (2001)
3. Jones, D.K., Griffin, L.D., Alexander, D.C., Catani, M., Horsfield, M.A., Howard, R., Williams, S.C.R.: Spatial normalization and averaging of diffusion tensor MRI data sets. NeuroImage 17, 592–617 (2002)
4. Park, H.J., Kubicki, M., Shenton, M.E., Guimond, A., McCarley, R., Maier, S.E., Kikinis, R., Jolesz, F.A., Westin, C.F.: Spatial normalization of diffusion tensor MRI using multiple channels. NeuroImage 20, 1995–2009 (2003)
5. Ziyan, U., Sabuncu, M.R., O'Donnell, L.J., Westin, C.-F.: Nonlinear registration of diffusion MR images based on fiber bundles. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 351–358. Springer, Heidelberg (2007)
6. Cao, Y., Miller, M., Mori, S., Winslow, R.L., Younes, L.: Diffeomorphic matching of diffusion tensor images. In: Proc. CVPRW 2006, p. 67 (2006)
7. Yeo, B., Vercauteren, T., Fillard, P., Pennec, X., Golland, P., Ayache, N., Clatz, O.: DTI registration with exact finite-strain differential. In: ISBI, pp. 700–703 (2008)
8. Jian, B., Vemuri, B.C., Özarslan, E., Carney, P.R., Mareci, T.H.: A novel tensor distribution model for the DWMR signal. NeuroImage 37, 164–176 (2007)
9. Jian, B., Vemuri, B.C.: A unified computational framework for deconvolution to reconstruct multiple fibers from DWMRI. IEEE TMI 26, 1464–1471 (2007)
10. Barmpoutis, A., Jian, B., Vemuri, B.C., Shepherd, T.M.: Symmetric positive 4th order tensors & their estimation from diffusion weighted MRI. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 308–319. Springer, Heidelberg (2007)
11. Özarslan, E., Mareci, T.H.: Generalized diffusion tensor imaging and analytical relationships between diffusion tensor imaging and high angular resolution diffusion imaging. Magn. Reson. Med. 50, 955–965 (2003)
12. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Apparent diffusion coefficients from high angular resolution diffusion imaging: Estimation and applications. Magn. Reson. Med. 56, 395–410 (2006)
13. Barmpoutis, A., Vemuri, B.C., Forder, J.R.: Registration of high angular resolution diffusion MRI images using 4th order tensors. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 908–915. Springer, Heidelberg (2007)
14. Chiang, M.-C., Leow, A.D., Klunder, A.D., Dutton, R.A., Barysheva, M., Rose, S.E., McMahon, K.L., de Zubicaray, G.I., Toga, A.W., Thompson, P.M.: Fluid registration of DT images using information theory. IEEE TMI 27, 442–456 (2008)
15. Jian, B., Vemuri, B.C.: A robust algorithm for point set registration using mixture of gaussian. In: ICCV, pp. 1246–1251 (2005)
16. Nocedal, J., Wright, S.: Numerical Optimization. Springer, Heidelberg (1999)
17. Soderman, O., Jonsson, B.: Restricted diffusion in cylindrical geometry. J. Magn. Reson. A 117, 94–97 (1995)
18. Zhang, H., Yushkevich, P.A., Gee, J.C.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 172–179. Springer, Heidelberg (2005)

Toward Real-Time Simulation of Blood-Coil Interaction during Aneurysm Embolization

Yiyi Wei[1,2], Stéphane Cotin[1], Le Fang[3], Jérémie Allard[1], Chunhong Pan[2], and Songde Ma[2]

[1] INRIA Lille - North Europe, 59650 Villeneuve d'Ascq, France
[2] LIAMA CASIA, 100190 Beijing, China
[3] LMFA, Ecole Centrale de Lyon, 69130, Ecully, France

Abstract. Over the last decade, remarkable progress has been made in the field of endovascular treatment of aneurysms. Technological advances continue to enable a growing number of patients with cerebral aneurysms to be treated with a variety of endovascular strategies, essentially using detachable platinum coils. Yet, coil embolization remains a very complex medical procedure for which careful planning must be combined with advanced technical skills in order to be successful.

In this paper we propose a method for computing the complex blood flow patterns that take place within the aneurysm, and for simulating the interaction of coils with this flow. This interaction is twofold, first involving the impact of the flow on the coil during the initial stages of its deployment, and second concerning the decrease of blood velocity within the aneurysm, as a consequence of coil packing. We also propose an approach to achieve real-time computation of coil-flow bilateral influence, necessary for interactive simulation. This in turns allows to dynamically plan coil embolization for two key steps of the procedure: choice and placement of the first coils, and assessment of the number of coils necessary to reduce aneurysmal blood velocity and wall pressure.

1 Introduction

Detachable coil embolization is a recent interventional technique for treating aneurysms and other blood vessel malformations in the brain and other parts of the body. The procedure uses the vascular network to reach the diseased vessel, starting with the insertion of a catheter (a long, thin and flexible tube) into the femoral artery. This catheter is then advanced through the arterial system until the aneurysm location is reached. Once in position, the physician places several coils through a micro-catheter into the aneurysm. The presence of coils reduces blood flow and wall pressure within the aneurysm, thus creating a favorable hemodynamic environment for thrombus embolization. The formation of a blood clot around the coil blocks off the aneurysm, thus considerably reducing the risk of rupture. Although coil embolization is less invasive than open surgery, such procedures are very difficult to perform and require careful planning and a long experience to minimize the risks for the patient. Yet, even in the case of a

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 198–205, 2009.

successfully performed procedure, the choice of the coil (shape, length, diameter) plays a key role in the long term success of the procedure. In this context, the development of an interactive planning system, allowing interventional radiologists to select different coils and test their behavior in a patient-specific environment, could make a difference. This requires to not only model the behavior of a coil in a patient-specific model of the aneurysm, but also to compute the interaction between a coil and the complex flow occurring within the aneurysm.

Previous Work. Blood flow dynamics is starting to play an increasingly important role in the assessment of vascular pathologies, as well as in the evaluation of pre- and post-operative status. While angiography has been an integral part of interventional radiology procedures for years, it is only recently that detailed analysis of blood flow patterns has been studied as a mean to assess complex procedures, such as coil deployment. A few studies have focused on aneurysm-related haemodynamics before and after endovascular coil embolization. Groden *et al.* [1] constructed a simple geometrical model to approximate an actual aneurysm, and evaluated the impact of different levels of coil packing on the flow and wall pressure by solving Navier-Stokes equations. Kakalis *et al.* [2] employed patient-specific data to get more realistic flow patterns, and modeled the coiled part, from a static point of view, as a porous medium. As these studies aimed at accurate Computational Fluid Dynamics simulation, they rely on commercial software, and the computation times (dozens of hours in general) are incompatible with interactive simulation or even clinical practice. Generally speaking, accuracy and efficiency are two significant pursuits in numerical calculation, but unfortunately always contradictory.

In the field of Computer Graphics, the main objective is to obtain efficient computations, yet still be able to capture the main structures of fluid motion. Stable fluids approach [3] was a significant milestone, as it brought in fluid advection and the Helmholtz-Hodge decomposition to ensure the mass conservation law. However, this approach relies on a discretization of the Eulerian space by a regular grid, thus making it inappropriate for simulations requiring irregular boundaries, as it is the case in medical applications. Recently, the Discrete Exterior Calculus (DEC) [4] method settled this problem, providing a mean to handle arbitrary meshes. This approach can be considered as an extension of Finite Difference Method to an arbitrary discretization of space. By preserving circulation at a discrete level it also guarantees a certain level of accuracy.

Regarding fluid/solids interaction, various models have been put forth. For instance, Carlson *et al.* [5] solved this problem by considering rigid objects as if they were made of fluid, and the rigidity was maintained by constraining the velocity field in the object region. Robinson-Mosher *et al.* [6] proposed a two-way solid/fluid coupling method to compute mutual effects separately at each time step. The effect of the solid on the fluid is modeled by taking solid velocity as a boundary condition, while the effect of the fluid on the solid is determined by integrating the fluid pressure along the solid boundary. Considering the dimension of coils and their nest-shaped distribution in an aneurysm, a different computational strategy is required in our case.

Finally, previous work in the area of real-time simulation of interventional radiology procedures mainly focused on training rather than planning (besides the work of [7]) and has been limited to the simulation of flexible devices without considering fluid interaction (see [8], or [9] for instance).

In this paper, we present a novel technique for accurately computing (eventually in real-time) the flow of blood within an aneurysm, as well as the interaction between blood and coils. We rely on the Discrete Exterior Calculus method to obtain an ideal trade-off between accuracy and computational efficiency. We also propose to pre-compute the flow velocity field over a cardiac cycle to achieve real-time computation of the interaction between coils and blood flow. Our results show that our approach can describe the influence between coils and blood flow.

2 Modeling Interactions between Blood Flow and Coils

In this section, we first describe the computational approach for computing blood flow in and around an aneurysm (section 2.1). To model the impact of the coil onto the flow (as a change of flow pattern and a decrease of velocity) we then introduce the notion of porous media, where coils are described from a statistical point of view, translating the idea that, after deployment, coils are randomly distributed in the aneurysm. We also show how the reverse effect, *i.e.*, the drag force applied onto the coil due to blood velocity, can be computed (section 2.2). This is particularly important during the first stage of coil deployment.

2.1 Blood Motion

Porous Media Model. We divide the fluid domain $\mathscr{D}$ (2D or 3D) into 2 sub-domains, a coil-free and a coil-filled subdomain. Blood motion in both sub-domains is described by a Navier-Stokes equation of Brinkmann type:

$$\begin{gathered}\frac{\partial(\varphi\rho\boldsymbol{u})}{\partial t}+\rho(\boldsymbol{u}\cdot\nabla)(\varphi\boldsymbol{u})=-\varphi\nabla p+\mu\triangle(\varphi\boldsymbol{u})-\frac{\varphi^2\mu}{k}\boldsymbol{u}-\frac{\varphi^3 C_D\rho}{\sqrt{k}}\boldsymbol{u}\left|\boldsymbol{u}\right| \\ \nabla(\varphi\boldsymbol{u})=0\end{gathered} \tag{1}$$

where $\boldsymbol{u}$ is the velocity of the fluid with density ρ and viscosity μ, and p is the pressure. The porosity φ and the permeability k are constitutive characteristics of the porous media, and C_D is the drag factor. The porosity φ describes the volume ratio of fluid to coil-filled subdomain, $\varphi = 1 - \frac{V_{coil}}{V_{an}}$, where V_{coil} is the accumulated volume of all coils, and V_{an} is the volume of the aneurysm. The permeability k measures the fluid conductivity through porous media, $k = \frac{\varphi^3}{cS^2}$, where c is the Kozeny coefficient (for cylinders, $c = 2$), and S is the ratio of the surface area of all coils to the volume of the aneurysm. The drag factor C_D can be derived from the computation of a local Reynolds number. In order to get rid of the pressure term, we take the curl of (1). As density and porosity remain constant in each subdomain, we have

$$\begin{gathered}\frac{\partial\boldsymbol{\omega}}{\partial t}+\mathscr{L}_{\boldsymbol{u}}\boldsymbol{\omega}=\frac{\mu}{\rho}\triangle\boldsymbol{\omega}-\frac{\mu\varphi}{\rho k}\boldsymbol{\omega}-\frac{\varphi^2 C_D}{\sqrt{k}}\nabla\times\boldsymbol{b} \\ \nabla(\varphi\boldsymbol{u})=0\quad\boldsymbol{\omega}=\nabla\times\boldsymbol{u}\quad\boldsymbol{b}=\boldsymbol{u}\left|\boldsymbol{u}\right|\end{gathered} \tag{2}$$

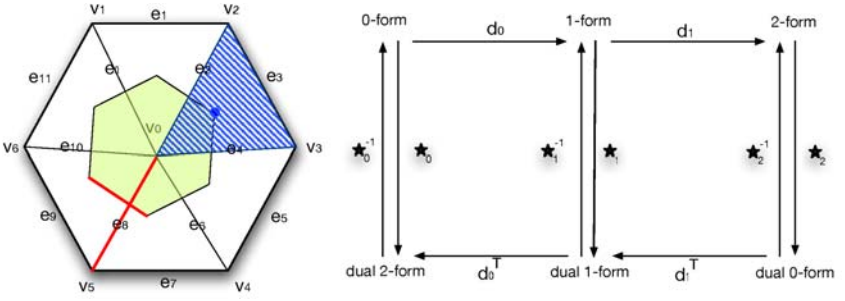


Fig. 1. (Left): two-dimensional mesh and its dual mesh, (right): transition graph between variables at primal and dual primitives, and operators describing the transitions

where $\boldsymbol{\omega}$ is the vorticity, $\mathscr{L}_{\boldsymbol{u}}\boldsymbol{\omega}$ is the Lie derivative, equal in our case to $\boldsymbol{u}\cdot\nabla\boldsymbol{\omega} - \boldsymbol{\omega}\cdot\nabla\boldsymbol{u}$, and $\boldsymbol{b}$ is the porous term.

Numerical Solution. As mentioned previously, the complex shape of aneurysms requires an unstructured grid to describe the geometry. We rely on the DEC method to provide a numerical framework for solving the fluid equations, by discretizing the space as a simplicial complex, and computing its dual complex. State variables are defined as discrete forms, *i.e.*, integral values over elements of these two meshes, complying with conservation laws at a discrete level, which is a key point to get accurate and stable results. Since the mesh is static (Eulerian approach), computation efficiency is obtained by pre-computing several vector calculus operators such as gradient and Laplace, which are defined using basic topological and geometrical operations (see [4] or [10] for details).

Discretization. In the following part, we limit our descriptions and illustrations to the 2D case for simplicity, but definitions in 3D are similar. The domain $\mathscr{D}$ is discretized as an oriented simplicial complex, *i.e.*, a triangulation, referred to as primal mesh. We denote the vertex set $V = \{v_i\}$, the edge set $E = \{e_{ij}\}$, and the triangle set $T = \{t_{ijk}\}(0 \leqslant i, j, k \leqslant |V|)$.

The dual mesh of a triangulation is constructed as follows: dual vertices correspond to the circumcenters of primal triangles, dual edges link dual vertices located on neighbor triangles, and dual faces are defined as Voronoi cells. More generally, a dual $(n-p)$-cell is associated to a corresponding p-simplex ($p = 0, 1, 2, n = 2$ for 2D mesh) as depicted in Fig. 1. Problem variables are defined as discrete p-forms, *i.e.*, scalars associated to p dimension primitives (p-cells) of either the primal or dual mesh.

Operators. All the vector calculus operators involved in our computation can be derived from two types of fundamental operators: the discrete differentials d and the hodge stars $\star$. The former (d) maps between discrete forms on the same mesh, represented by the transpose of the signed incidence matrix, while the latter ($\star$) transforms discrete forms from one mesh to the other, represented by a diagonal matrix whose element equals to the volume ratio between the corresponding dual and primal elements (see Fig. 1).

$C \leftarrow d_0^t \star_1$ // Load mesh and compute the operators
$L \leftarrow d_0^t \star_1 d_0$
$L' \leftarrow \text{LagrangeMultiplier}(L)$
loop // Time stepping $\triangle t$
  $\hat{\mathbf{c}} \leftarrow \text{backtrackTriangleCenters}(\triangle t)$ // advect vorticities
  $\hat{\mathbf{v}} \leftarrow \text{interpolateVelocity}(\hat{\mathbf{c}})$
  for each *dual face* f
    $\boldsymbol{\Omega}_f \leftarrow 0$
    for each *dual edge* (i, j) *on the boundary of* f
      $\boldsymbol{\Omega}_f \leftarrow \boldsymbol{\Omega}_f + \frac{1}{2}(\hat{\mathbf{v}}_i + \hat{\mathbf{v}}_j)(\hat{\mathbf{c}}_i - \hat{\mathbf{c}}_j)$
  for each *dual face* f *in the coil-filled subdomain* // add porous terms
    $\boldsymbol{\Omega}_f \leftarrow \frac{\rho k}{\rho k + \mu\varphi\triangle t}\boldsymbol{\Omega}_f + \frac{\varphi^2 C_D}{\sqrt{k}}\triangle t\ (C \times \mathbf{B})_f$
  $\boldsymbol{\Psi} \leftarrow \text{linearSolver}(\ (\star_0 - \mu L \triangle t)\boldsymbol{\Psi} = \boldsymbol{\Omega})\ )$ // add viscous term
  $\boldsymbol{\Omega} \leftarrow \star_0 \boldsymbol{\Psi}$
  $\tau \leftarrow \text{setBoundaryConditions}(t)$ // recover flux from vorticity
  $(\boldsymbol{\Phi} \mid \lambda) \leftarrow \text{linearSolver}(\ L'(\boldsymbol{\Phi} \mid \lambda) = (\boldsymbol{\Omega} \mid \tau)\ )$
  $\mathbf{U} = d_0 \boldsymbol{\Phi}$
  foreach *triangle* f **do** $\mathbf{v}_f \leftarrow A_f^{-1}\mathbf{U}_f$ // velocity at triangle centers
  $\mathbf{B} \leftarrow \text{integrate}(\mathbf{v})$

Fig. 2. Pseudocode of Fluid Computation

Solving Fluid Equations. In DEC, velocity is described as flux, *i.e.*, the mass of fluid passing through each edge per unit time. Thus it's a discrete 1-form $\mathbf{U}$, represented as a vector of size $|E|$. Similarly, the porous term $\boldsymbol{b}$ is defined as a discrete 1-form $\mathbf{B}$. In continuum fluid dynamics, vorticity measures the local angular rate of rotation, defined as the circulation per unit area at a point. Accordingly, we describe discrete vorticity $\boldsymbol{\Omega}$ through the integral over faces, and a natural choice is the dual face, in that we can easily get these values from $\mathbf{U}$ as follows: $\star_1$ transforms $\mathbf{U}$ on primal edges to $\star_1\mathbf{U}$ on dual edges, and then d_0^t sums it on each dual faces by accumulating $\star_1\mathbf{U}$ on all incident dual edges, *i.e.*, $\boldsymbol{\Omega} = d_0^t \star_1 \mathbf{U}$, a dual 2-form. This explains how $\nabla\times$ is built by the basic operators. Following similar principles, all operators (grad $\nabla\cdot$, curl $\nabla\times$, Laplace $\triangle$) used in (2) can be constructed from d and $\star$: $\nabla\cdot = d_1$, $\nabla\times = d_0^t\star_1$, $\triangle = d_0^t \star_1 d_0$.

The vorticity-based equation (2), simply speaking, describes the idea that the local spin is pushed forward along the direction of the velocity. This is consistent with Kelvin's circulation theorem: the circulation around a closed curve moving with the fluid remains constant with time [10]. In our approach, the discrete vorticity is conserved by extending Kelvin's theorem to the discrete level: the circulation around the loop of each dual face's boundary keeps constant as the loop is advected by fluid flow. We implement this by *backtracking* all the loops into the previous time step [3]. Once the vorticity have been computed, the flux is recovered by solving the Laplace equation, taking into account the varying boundary conditions. Finally, the velocity at each triangle centers is computed using the property that there is a unique vector at each dual vertex whose projection along the incident dual edges is consistent with the flux of the corresponding

primal edges. This defines a projection matrix A_f for each triangle f. The final computation process is shown in Fig. 2.

2.2 Coil Motion

In the existing simulations of aneurysm embolization, the interactive force between blood and coil was only studied for the blood, while the reacting force on coils was ignored. In fact, the last term of (1) is a description of the interactive force, but treated as an averaged quantity. When computing the reaction on the coil, we apply its local version, which is the drag force of flow over a cylinder:

$$F_D = \frac{1}{2} C_D \rho \boldsymbol{u}_\perp |\boldsymbol{u}_\perp| A dh, \tag{3}$$

where $\boldsymbol{u}_\perp$ is the velocity orthogonal to the coil, A is the cross-sectional area of the coil, dh is the length of the coil section. The velocity parallel to the coil is neglected, since it only produces shear force on the coil, which is insignificant compared to the drag force. Hence, the reacting force on the coil only depends on local fluid velocity. The coil model is based on the work of Dequidt *et. al.* [7] where coils are modeled as a series of serially-linked beam elements. The deformation of the structure is computed using a finite element approach which can be optimized for real-time computation by taking advantage of the structure (tri-diagonal band) of its stiffness matrix.

3 Real-Time Simulation

In this section, we show how the two previously described models can be combined and used in a real-time simulation. For this we assume the simulation is performed over a series of identical cardiac cycles. Periodically time-varying boundary conditions are set at the inlet and outlet vessels around the aneurysm for a duration covering a complete cardiac cycle. Using the method described in section 2.1 we compute and store the velocity field for multiple time steps within one cardiac period. This process can be done for different densities of coil in the aneurysm. This database of velocity fields can then be used to interpolate the velocity at the position of coil segments and apply appropriate drag forces. It can also provide real-time feedback, at any step of the embolization, about blood velocities inside the aneurysm. It should also be noted that most of the computation time of our method comes from solving two linear sparse systems of equations. Many numerical techniques can be used to improve the efficiency of this process. When dealing with relatively small systems a direct inversion of the matrices is often the best approach. The inversion of these matrices can be performed only once, during initialization. This ensures real-time or near real-time computation (which is typically the case when dealing with two-dimensional problems). For larger systems, we rely on the offline computation of operators and a pre-factorization of the sparse linear systems. This is however not always sufficient to achieve real-time computation when high accuracy is needed.

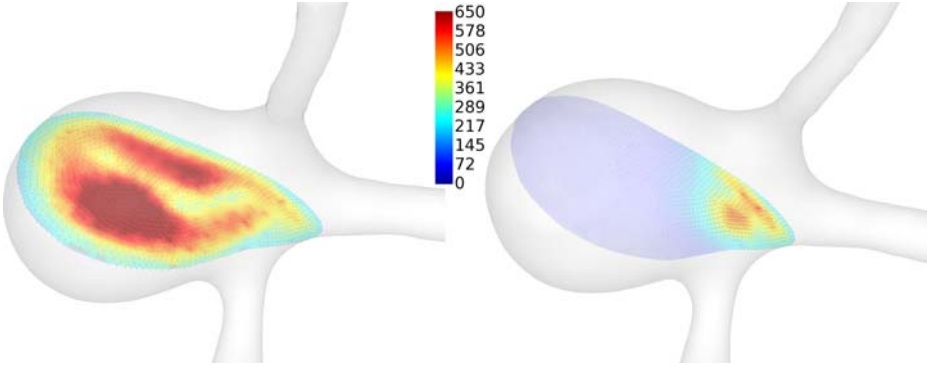


Fig. 3. (Left): velocity magnitude in aneurysm with no coil; (Right): with 16.8% of the volume filled with coils. Aneurysm model composed of of 18K tetrahedral elements.

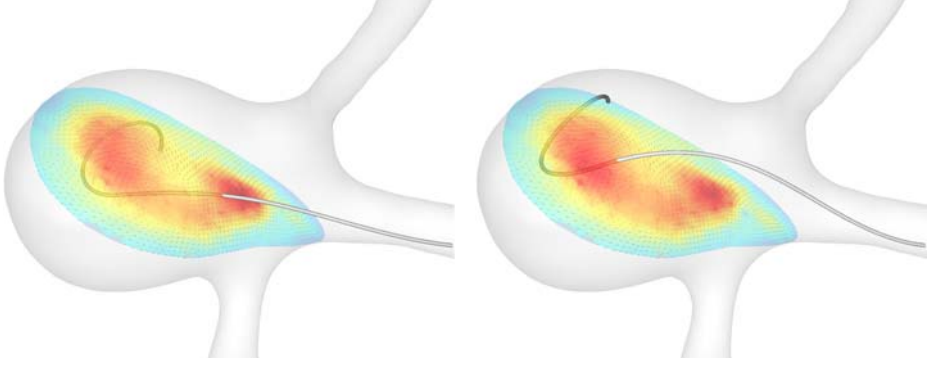


Fig. 4. (Left): Coil shape when the effect of blood flow is not taken into consideration; (Right): Resulting shape when considering flow forces on the coil

4 Results

Our simulations are performed on an aneurysm of particular interest, with a large sac of volume $8.976 \cdot 10^{-7}\text{m}^3$ and a wide neck of dimension $8.2 \cdot 10^{-3}\text{m}$. Such aneurysms are difficult to treat, as coils might be pushed out by the blood flow during deployment. The aneurysm geometry is obtained from CT images, and then discretized into a tetrahedral mesh.

The parameters of our model were set as follows: blood density: 1069kg/m^3, blood viscosity: 0.0035kg/ms, drag factor C_D: 2.2, coil radius: $0.36 \cdot 10^{-3}\text{m}$. Fig. 3 shows the velocity magnitude contours before and after placement of coils of total length 0.3m ($\varphi = 83.2\%$). The decrease of velocity magnitude is obvious and in accordance with recent results [2]. But we need only 25s of simulation for a cardiac cycle of 1s (using a mesh of 18K tetrahedra and a time step of 0.002s). Using a pre-computed time-varying velocity field for a complete cardiac cycle, we simulated the influence of the flow on a typical coil (radius of $0.36 \cdot 10^{-3}\text{m}$,

Young modulus of 10^9Pa, length of 10cm) discretized using 100 elements. The resulting simulation, illustrated in Fig. 4, was performed in real-time.

5 Conclusion

In this paper we present a method for efficiently and accurately computing blood flow in aneurysms. More importantly we introduce parameters to account for the presence of coils in the aneurysm in order to model their impact on the flow. We also model the reciprocal effect, i.e. the impact of the flow onto the coil. Both aspects are essential in the context of coil embolization planning. Our method is significantly faster than previous approaches while providing similar results. We have assessed our computation on an aneurysm presenting interesting clinical characteristics.

Regarding future directions, we acknowledge that further validation is required, however experimental data on flow patterns in aneurysms is challenging to obtain. We also want to investigate more deeply various computational strategies to obtain real-time (or near real-time) computation by using more advanced numerical schemes. Also, using the DEC method, several steps of our computation depend only on topological neighbors. This could be leveraged to provide a parallel implementation on GPU.

References

1. Groden, C., Laudan, J., Gatchell, S., Zeumer, H.: Three-dimensional pulsatile flow simulation before and after endovascular coil embolization of a terminal cerebral aneurysm. J. Cereb. Blood Flow Metab. 21(12), 1464–1471 (2001)
2. Kakalis, N., Mitsos, A.P., Byrne, J.V., Ventikos, Y.: The haemodynamics of endovascular aneurysm treatment: a computational modelling approach for estimating the influence of multiple coil deployment. IEEE Trans. Med. Imaging 27(6), 814–824 (2008)
3. Stam, J.: Stable fluids. In: Proceedings of SIGGRAPH 1999, pp. 121–128. ACM Press, New York (1999)
4. Desbrun, M., Kanso, E., Tong, Y.: Discrete differential forms for computational modeling. In: ACM SIGGRAPH Courses, ch. 7 (2005)
5. Carlson, M., Mucha, P.J., Turk, G.: Rigid fluid: Animating the interplay between rigid bodies and fluid. ACM Trans. Graph, 377–384 (2004)
6. Robinson-Mosher, A., Shinar, T., Gretarsson, J., Su, J., Fedkiw, R.: Two-way coupling of fluids to rigid and deformable solids and shells. ACM Trans. Graph. 27(3), 1–9 (2008)
7. Dequidt, J., Marchal, M., Duriez, C., Kerien, E., Cotin, S.: Interactive simulation of embolization coils: Modeling and experimental validation. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 695–702. Springer, Heidelberg (2008)
8. Alderliesten, T.: Simulation of Minimally-Invasive Vascular Interventions for Training Purposes. PhD dissertation, Utrecht University (2004)
9. Duriez, C., Cotin, S., Lenoir, J., Neumann, P.F.: New approaches to catheter navigation for interventional radiology simulation. Computer Aided Surgery 11, 300–308 (2006)
10. Elcott, S., Tong, Y., Kanso, E., Schroder, P., Desbrun, M.: Stable, circulation-preserving, simplicial fluids. In: ACM SIGGRAPH 2006, Courses, pp. 60–68 (2006)

A Dynamical Shape Prior for LV Segmentation from RT3D Echocardiography*

Yun Zhu[1], Xenophon Papademetris[1,2], Albert J. Sinusas[2], and James S. Duncan[1,2]

[1] Department of Biomedical Engineering, Yale University, USA
yun.zhu@yale.edu
[2] Department of Diagnostic Radiology, Yale University, USA

Abstract. Real-time three-dimensional (RT3D) echocardiography is the newest generation of three-dimensional (3-D) echocardiography. Segmentation of RT3D echocardiographic images is essential for determining many important diagnostic parameters. In cardiac imaging, since the heart is a moving organ, prior knowledge regarding its shape and motion patterns becomes an important component for the segmentation task. However, most previous cardiac models are either static models (SM), which neglect the temporal coherence of a cardiac sequence or generic dynamical models (GDM), which neglect the inter-subject variability of cardiac motion. In this paper, we present a subject-specific dynamical model (SSDM) which simultaneously handles inter-subject variability and cardiac dynamics (intra-subject variability). It can progressively predict the shape and motion patterns of a new sequence at the current frame based on the shapes observed in the past frames. The incorporation of this SSDM into the segmentation process is formulated in a recursive Bayesian framework. This results in a segmentation of each frame based on the intensity information of the current frame, as well as on the prediction from the previous frames. Quantitative results on 15 RT3D echocardiographic sequences show that automatic segmentation with SSDM is superior to that of either SM or GDM, and is comparable to manual segmentation.

1 Introduction

Real-time three-dimensional (RT3D) echocardiography is a new imaging modality that can capture the complex three-dimensional (3-D) shape and motion of the heart *in vivo*. To fully take advantage of the information offered by RT3D echocardiography, a robust and accurate automatic segmentation tool for tracking the dynamic shape of the heart is indispensable for quantitative analysis of cardiac function. Unfortunately, automatic segmentation of RT3D echocardiography is challenging and depends on image quality.

The use of shape and temporal priors has proven effective for segmenting images with missing and misleading image information [1]. The most widely used

* This work is supported by the grant 5R01HL082640-03.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 206–213, 2009.

shape model is probably the Active Shape Model (ASM), which uses Principal Components Analysis (PCA) to describe the average shape and the most characteristic shape variations of a set of training shapes. However, ASM is a static model (SM) because it supplies a prior just for shape, but not for the motion of that shape. Temporal models can take a number of forms. The simplest form is one that insists on temporal coherence and smoothness. More complicated forms approximate cardiac motion using a parametric model. However, these temporal models do not include any prior knowledge of the shape.

To combine shape and temporal priors, researchers have proposed spatial-temporal statistical models. For example, Mitchell *et al.* extended ASM to the 2-D Active Appearance Motion Model (AAMM) [2], which includes both motion and appearance information. It is difficult, however, to extend 2-D AAMM to the segmentation of a full 3-D cardiac sequence because of the high dimensionality involved. Perperidis *et al.* constructed a 4-D atlas using two separate models, which accounted for inter-subject variability and cardiac temporal dynamics (intra-subject variability), respectively [3]. While these two models were successfully applied to the classification of cardiac images from normal volunteers and patients with hypertrophic cardiomyopathy, they are not related, making them suboptimal for left ventricular (LV) segmentation.

Dynamical shape model is a recently proposed spatial-temporal statistical model. It performs sequential segmentation using the cardiac dynamics learned from a set of training samples. It is highly flexible, and can be applied to segment a full 3-D sequence. For example, Jacob *et al.* proposed a second-order autoregressive model to approximate cardiac dynamics [4]. Sun *et al.* proposed learning cardiac dynamics using a second-order nonlinear model [5]. While these models are superior to SM, they are time homogeneous[1] and therefore inadequate for describing complex shape deformations, such as cardiac dynamics. In addition, since they supply a uniform model to all sequences, they ignore the subject variations in motion patterns. This makes them generic dynamical models (GDM).

In this paper, we present a subject-specific dynamical model (SSDM) to simultaneously account for the subject-specific variations in cardiac shape as well as inhomogeneous motion patterns. To build this SSDM, we need to differentiate two factors that cause cardiac shape variability. One is the inter-subject variability, and the other is temporal dynamics caused by cardiac deformation during a cardiac cycle, as shown in Figure 1. These two factors are interactive and cannot be separated into two independent statistical models. Because conventional PCA and Independent Component Analysis (ICA) can only focus on one factor at a time, we extend them to higher orders by utilizing Multilinear PCA (MPCA) [6] and Multilinear ICA (MICA) [7]. This allows us to decompose the training set and to describe the interaction of inter-subject variability and temporal dynamics. In addition, we design a dynamic prediction algorithm that can progressively

[1] A dynamical model is time-homogeneous if the conditional probability of state t given its previous states only depends on the time difference between those states, i.e. $\mathcal{P}\left(\mathbf{s}_t|\mathbf{s}_{t-1}, \mathbf{s}_{t-2}, ..., \mathbf{s}_{t-m}\right) = \mathcal{P}\left(\mathbf{s}_{t+n}|\mathbf{s}_{t+n-1}, \mathbf{s}_{t+n-2}, ..., \mathbf{s}_{t+n-m}\right)$, for all n and m.

identify the subject vector associated with a new cardiac sequence and use this subject vector to predict the subject-specific dynamics from the segmentations observed in the previous frames. We formulate the integration of this SSDM into a recursive Bayesian framework. This framework models the evolution of the endocardial (ENDO) and epicardial (EPI) surfaces driven by both intensity information from the current frame as well as the dynamical shape prior inferred from the past segmentations, based on the knowledge learned from the training set.

2 Method

2.1 The Construction of SSDM

Shape Alignment. Magnetic Resonance (MR) images have higher a spatial resolution and signal-to-noise (SNR) ratio than RT3D echocardiography, and therefore are appropriate for building the SSDM. We acquired 32 sequences of electrocardiography (ECG)-gated canine short-axis MR images, with 16 temporal frames per sequence. The in-plane resolution was 1.6 mm, and the slice thickness was 5 mm. The ENDO and EPI surfaces were manually outlined by an experienced cardiologist using the BioImage Suite software [8]. We first extracted 153 landmarks on the ENDO surface and 109 landmarks on the EPI surface in the first frame of the first sequence. Then we propagated this set of landmarks to all frames in each sequence by mapping those frames to the first frame of the first sequence using inter- and intra-subject registrations, as shown in Figure 1. We used an affine transform to account for the global shape difference, as well as a shape-based non-rigid transform, as described in [9], to accommodate the detailed shape differences. Thus, we obtained 262 landmarks for each frame.

Shape Decomposition. In this paper, we use MPCA and MICA to decompose cardiac shapes (see [6,7] for an overview of MPCA and MICA). Here we denote the aligned cardiac shapes as third-order tensor $\mathcal{S} \in \mathbb{R}^{I \times J \times K}$, where $I = 32$ is the number of subjects, $J = 16$ is the number of frames within a sequence, and $K = 262 \times 3 = 786$ is the dimension of landmark vectors. By applying MPCA to tensor $\mathcal{S}$, we have

$$\mathcal{S} \approx \mathcal{Z} \times_1 \mathbf{U}^{\text{subject}} \times_2 \mathbf{U}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}} \tag{1}$$

where $\mathcal{Z} \in \mathbb{R}^{P \times Q \times R}$ is the core tensor, which represents the interaction of the subject, motion, and landmark subspaces. Matrices $\mathbf{U}^{\text{subject}} \in \mathbb{R}^{I \times P}$, $\mathbf{U}^{\text{motion}} \in \mathbb{R}^{J \times Q}$, and $\mathbf{U}^{\text{landmark}} \in \mathbb{R}^{K \times R}$ are the subject subspace, motion subspace, and landmark subspace respectively. Matrix $\mathbf{U}^{\text{subject}}$ contains row vectors $\mathbf{u}_i^{\text{subject}} \in \mathbb{R}^P$ $(1 \leq i \leq I)$ of coefficients for each person i, and matrix $\mathbf{U}^{\text{motion}}$ contains row vectors $\mathbf{u}_j^{\text{motion}} \in \mathbb{R}^Q$ $(1 \leq j \leq J)$ for frame j.

While it is reasonable to apply PCA in the subject subspace, it is inappropriate to use it in the motion subspace because the deformation of cardiac shapes does not have a Gaussian distribution. To handle this problem, we adopt ICA

in the motion subspace to obtain a set of independent modes in the motion subspace [10]. We rewrite Equation 1 as $\mathcal{S} \approx \mathcal{Z} \times_1 \mathbf{U}^{\text{subject}} \times_2 \mathbf{U}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}} = \mathcal{Z} \times_1 \mathbf{U}^{\text{subject}} \times_2 \mathbf{U}^{\text{motion}} \mathbf{W}^T \mathbf{W}^{-T} \times_3 \mathbf{U}^{\text{landmark}} = \tilde{\mathcal{Z}} \times_1 \mathbf{U}^{\text{subject}} \times_2 \tilde{\mathbf{U}}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}}$, where the core tensor $\tilde{\mathcal{Z}} = \mathcal{Z} \times_2 \mathbf{W}^{-T}$, the column vectors of $\tilde{\mathbf{U}}^{\text{motion}}$ are independent components of the motion subspace $\tilde{\mathbf{U}}^{\text{motion}}$.

We have two steps to reduce dimensions. First, we select the complete eigenvectors in the motion subspace, i.e. $Q = J$, and perform MPCA in the subject and landmark subspace to find the optimal P and R such that the approximation keeps more than 98% of the original energy. Second, we fix P and R, and perform MICA in the motion subspace to find the modes that correspond to significant shape variations. In practice, we reduced $I = 32$ to $P = 5$ and $K = 786$ to $R = 11$ in the first step, and further reduced $J = 16$ to $Q = 3$ in the second step.

Dynamic Prediction. Given the segmentation of a new cardiac sequence from frame 1 to $t-1$, we want to predict its segmentation in frame t. The idea is to first project the given segmentation from frame 1 to $t-1$ to the subject subspace to identify the subject vector associated with this sequence, and then to use this subject vector to predict the LV shape at frame t.

Let $\mathbf{s}_{1:t-1} = \{\mathbf{s}_1, \mathbf{s}_2, ..., \mathbf{s}_{t-1}\}$ denote the observed segmentation of a new cardiac sequence. We predict the segmentation at frame t using two steps: projection and prediction. In the projection step, we estimate the subject vector associated with this new sequence as $\hat{\mathbf{u}}^{\text{subject}} = \mathbf{s}_{1:t-1} \cdot \mathbf{T}_{(1)}^{-1}$, where $\mathbf{T}_{(1)}$ is the mode-1 unfolding of tensor $\mathcal{T} = \tilde{\mathcal{Z}} \times_2 \tilde{\mathbf{u}}_{1:t-1}^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}}$. In the prediction step, we use the subject vector estimated in the projection step to predict the segmentation at frame t as $\mathbf{s}_t^* = \tilde{\mathcal{Z}} \times_1 \hat{\mathbf{u}}^{\text{subject}} \times_2 \tilde{\mathbf{u}}_t^{\text{motion}} \times_3 \mathbf{U}^{\text{landmark}}$.

As mentioned above, we used MR sequences to build the SSDM, and then used this SSDM to predict the cardiac dynamics of a new RT3D echocardiographic sequence. However, this sequence may have a different length of cardiac cycle from the MR sequences. To handle this problem, we first align the end-diastolic

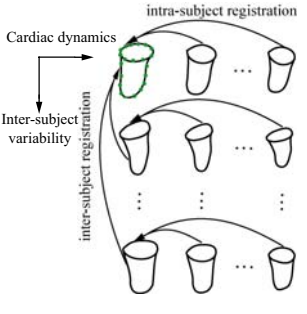


Fig. 1. The interaction of cardiac dynamics (intra-subject variability) and inter-subject variability

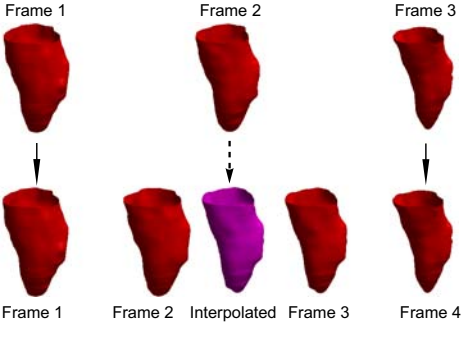


Fig. 2. Temporal interpolation to generate time frames from sequence 2 which correspond to those of sequence 1

and end-systolic frames using ECG signals, and then use linear interpolation to generate frames that correspond to the RT3D echocardiographic frames, as shown in Figure 2.

2.2 General Formulation

Assume that we are given a cardiac sequence $I_{1:t} = \{I_1, I_2, ..., I_t\}$, and let $\mathbf{s}_t = \left\{\mathbf{s}_t^+, \mathbf{s}_t^-\right\}$ be the segmentation at frame t, where $\mathbf{s}_t^+$ is the ENDO surface, and $\mathbf{s}_t^-$ is the EPI surface. The problem of segmenting the current frame t can be addressed by maximizing the conditional probability

$$\begin{aligned} &\mathcal{P}\left(\mathbf{s}_t|I_{1:t}\right) \propto \mathcal{P}\left(I_t|\mathbf{s}_t, I_{1:t-1}\right)\mathcal{P}\left(\mathbf{s}_t|I_{1:t-1}\right) \\ &= \mathcal{P}\left(I_t|\mathbf{s}_t, I_{1:t-1}\right)\int \mathcal{P}\left(\mathbf{s}_t|\mathbf{s}_{1:t-1}\right)\mathcal{P}\left(\mathbf{s}_{1:t-1}|I_{1:t-1}\right)d\mathbf{s}_{1:t-1} \end{aligned} \tag{2}$$

In the following, we make two assumptions in order to lead to a computationally more feasible problem [11]. (1) The images $I_{1:t}$ are mutually independent, i.e. $\mathcal{P}\left(I_t|\mathbf{s}_t, I_{1:t-1}\right) = \mathcal{P}\left(I_t|\mathbf{s}_t\right)$. (2) The distributions of previous states are strongly peaked around the maxima of the respective distributions, i.e. $\mathcal{P}\left(\mathbf{s}_{1:t-1}|I_{1:t-1}\right) \approx \delta\left(\mathbf{s}_{1:t-1} - \hat{\mathbf{s}}_{1:t-1}\right)$, where $\hat{\mathbf{s}}_i = \arg\max \mathcal{P}\left(\mathbf{s}_i|I_{1:i}\right)$ and $\delta\left(\cdot\right)$ is Dirac delta function.

Thus, we have $\mathcal{P}\left(\mathbf{s}_t|I_{1:t}\right) \propto \underbrace{\mathcal{P}\left(I_t|\mathbf{s}_t\right)}_{\text{data adherence}} \underbrace{\mathcal{P}\left(\mathbf{s}_t|\hat{\mathbf{s}}_{1:t-1}\right)}_{\text{dynamical shape prior}}$. It is a recursive Bayesian formulation, where the ENDO and EPI contours are driven not only by the intensity information from the current frame, but also from the dynamical shape prior from the past frames.

2.3 Data Adherence

An entire cardiac image is partitioned by the ENDO and EPI contours into three regions: LV blood pool, LV myocardium, and background. The simplest intensity distribution for B-mode images is Rayleigh distribution [1]. However, Rayleigh distribution is only effective for fully-developed speckles. More complicated models, such as the Rice distribution and the K-distribution, have been proposed to account for the regular structure of scatters and low effective scatter density [1]. Unfortunately, the analytical complexity involved with these distributions is significant. In this paper, we utilize the Nakagami distribution [12], a simpler generalized distribution which can handle simultaneously the situations of regularly-spaced scatters and varying scatter densities. Thus, the intensity distribution for the LV blood pool and myocardium can be expressed as $\mathcal{P}_l\left(I;\mu_l,\omega_l\right) = \frac{2\mu_l^{\mu_l}}{\Gamma(\mu_l)\omega_l^{\mu_l}} I^{2\mu_l-1}\exp\left(-\frac{\mu_l}{\omega_l}I^2\right)$, where μ_l is the Nakagami parameter and ω_l is a scaling parameter. For $l = 1$, it models the intensity distribution in the LV blood pool. For $l = 2$, it models the intensity distribution in the LV myocardium.

Unlike the LV blood pool and myocardium, the background includes more than one tissue (e.g. RV blood pool, RV myocardium, and other tissues). Therefore, we use a mixture model and invoke the Expectation-Maximization (EM)

algorithm to fit the background histogram. Under the mixture model, the background distribution is given as $\mathcal{P}_3(I;\mu_3,\omega_3) = \sum_{k=1}^{M} \alpha_k \mathcal{P}_k(I;\mu_{3,k},\omega_{3,k})$, where M is the number of components, α_k is the mixture proportion of component k that satisfies $\sum_{k=1}^{M} \alpha_k = 1$, $\mu_{3,k}$ and $\omega_{3,k}$ are the parameters of its component distributions. In the experiments, we set $M = 2$.

Let Ω_1, Ω_2, and Ω_3 denote three regions: LV blood pool, LV myocardium, and background, respectively. Then, the data adherence term can be defined as follows

$$\log \mathcal{P}(I|\mathbf{s}) = \sum_{l=1}^{3} \int_{\Omega_l} \log \mathcal{P}_l(I;\mu_l,\omega_l)\, d\mathbf{x} \tag{3}$$

The maximization of Equation 3 can be interpreted as the propagation of $\mathbf{s} = \{\mathbf{s}^+, \mathbf{s}^-\}$ that maximizes the piecewise homogeneities.

2.4 Dynamical Shape Prior

As shown in Section 2.1, we predict the ENDO and EPI contours at frame t using the dynamic prediction algorithm. Thus, we define the dynamic prior as

$$\mathcal{P}(\mathbf{s}_t|\hat{\mathbf{s}}_{1:t-1}) \propto \exp\left\{-\alpha \left\|\mathbf{s}_t - \mathbf{s}_t^*\right\|^2 \Big/ 2\right\} \tag{4}$$

where α is a weighting parameter, and $\mathbf{s}_t^*$ is the predicted shape at frame t using the dynamic prediction algorithm described in Section 2.1. In the experiments, we found $1.5 \leq \alpha \leq 2.5$ is applicable to most of the data.

3 Results

Figure 3 represents the automatically segmented ENDO and EPI contours during ventricular systole. To further quantify the segmentation results, we asked two experts, blind to each other, to independently outline the ENDO and EPI contours of all of the frames of the image sequences. We then compared the manual results to the automatic results using three metrics: mean absolute distance (MAD), Hausdorff distance (HD), and the percentage of correctly segmented voxels (PTP). Let $A = \{\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_n\}$, $B = \{\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_m\}$, we define MAD $= \frac{1}{2}\left\{\frac{1}{n}\sum_{i=1}^{n} d(\mathbf{a}_i, B) + \frac{1}{m}\sum_{i=1}^{m} d(\mathbf{b}_i, A)\right\}$, HD $= \max\left(\max_i \{d(\mathbf{a}_i, B)\}, \max_j \{d(\mathbf{b}_j, A)\}\right)$, where $d(\mathbf{a}_i, B) = \min_j \|b_j - a_i\|$. Let Ω_a be the region enclosed by automatic segmentation, and Ω_m be the region enclosed by the manual segmentation, we define PTP $= \frac{\text{Volume}(\Omega_a \cap \Omega_m)}{\text{Volume}(\Omega_m)}$. While MAD represents the global disagreement between two contours, HD compares their local similarities.

Table 1 compares the segmentation results of ENDO and EPI contours from SM, GDM and SSDM. For ENDO contours, the automatic-manual MAD using

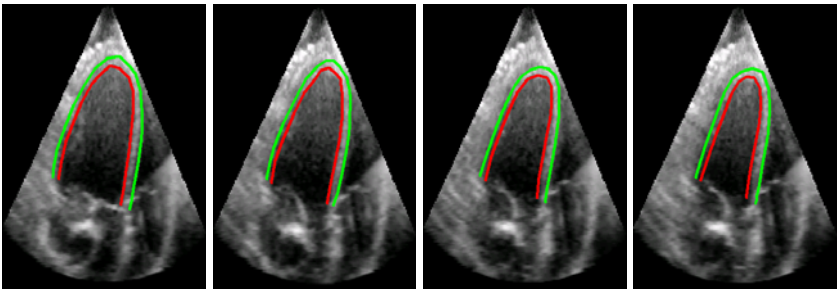


Fig. 3. The automatically segmented ENDO- and EPI contours during ventricular systole. Red: ENDO surface, Green: EPI surface.

Table 1. Comparison of automatic outline to two experts' outline of ENDO and EPI boundaries

		MAD (mm)	HD (mm)	PTP (%)
ENDO	automatic-manual (SSDM)	1.41 ± 0.40	2.53 ± 0.75	95.9 ± 1.24
	automatic-manual (GDM)	1.52 ± 0.46	3.25 ± 0.98	94.8 ± 1.56
	automatic-manual (SM)	2.33 ± 0.67	4.31 ± 1.26	93.1 ± 1.51
	manual-manual	1.37 ± 0.36	2.38 ± 0.65	95.8 ± 1.48
EPI	automatic-manual (SSDM)	1.74 ± 0.39	2.79 ± 0.97	94.5 ± 1.74
	automatic-manual (GDM)	1.77 ± 0.41	2.91 ± 0.95	93.6 ± 1.78
	automatic-manual (SM)	1.81 ± 0.65	3.18 ± 1.23	92.3 ± 1.91
	manual-manual	1.73 ± 0.51	2.83 ± 1.50	94.5 ± 1.77

SM was 0.92 mm larger than that obtained using SSDM, although the automatic-manual MAD using GDM was similar to that obtained using SSDM. This implies that both SSDM and GDM are able to capture the global deformation of cardiac shapes, while SM has tendency to get stuck in local minima because it does not provides a prediction in time. We also observed that the automatic-manual HD using SSDM was 0.72 mm larger than that obtained using GDM, and 1.78 mm larger than that with SM. This suggests that while GDM produced globally correct results, it failed to capture local shape deformations. Moreover, we observed that the performance of the automatic segmentation using SSDM was comparable to that of manual segmentation because similar MAD, HD, and PTP were produced. For EPI contours, in comparison with SM, the GDM improved the MAD by 0.04 mm, the HD by 0.13mm, and the PTP by 1.3%. When the SSDM was applied, the MAD was further improved by 0.03 mm, the HD by 0.12 mm, and the PTP by 0.9%. The improvement of EPI segmentation was less pronounced than that of ENDO segmentation. This is because the EPI surface does not move as much as the ENDO surface. Furthermore, we observed that the variability of manual-manual segmentation was smaller for the ENDO boundary than for the EPI boundary. This is probably because the EPI boundaries are more ambiguous for observers to detect, which was also the reason we used two observers, instead of a single one.

4 Conclusion

In this paper, we have presented a subject-specific dynamical model that utilized MPCA and MICA to simultaneously decompose the cardiac shape and motion in different subspaces. We then used a dynamic prediction algorithm to sequentially predict the dynamics of a new RT3D echocardiograpic sequence from the shapes observed in the past frames. Experiments on 15 sequences of echocardiographic data showed that automatic segmentation using the SSDM produced results that had an accuracy comparable to that obtained by manual segmentation. Future work would include the extension of this SSDM to human data and other modalities, and the development of an integrated framework that will combine cardiac segmentation and motion analysis for RT3D echocardiography.

References

1. Noble, J.A., Boukerroui, D.: Ultrasound image segmentation: A survey. IEEE TMI 25(8), 987–1010 (2006)
2. Bosch, J.G., Mitchell, S.C., Lelieveldt, B.P.F., Nijland, F., Kamp, O., Sonka, M., Reiber, J.H.C.: Automatic segmentation of echocardiographic sequences by active appearance motion models. IEEE TMI 21(11), 1374–1383 (2002)
3. Perperidis, D., Mohiaddin, R.H., Rueckert, D.: Construction of a 4D statistical atlas of the cardiac anatomy and its use in classification. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 402–410. Springer, Heidelberg (2005)
4. Jacob, G., Noble, J.A., Behrenbruch, C., Kelion, A., Banning, A.: A shape-space-based approach to tracking myocardial borders and quantifying regional left-ventricular function applied in echocardiography. IEEE TMI 21(3), 226–238 (2002)
5. Sun, W., Cetin, M., Chan, R., Reddy, V., Holmvang, G., Chandar, V., Willsky, A.: Segmenting and tracking the left ventricle by learning the dynamics in cardiac images. In: IPMI, pp. 553–565 (2005)
6. Vasilescu, M.A.O., Terzopoulos, D.: Multilinear analysis of image ensembles: TensorFaces. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 447–460. Springer, Heidelberg (2002)
7. Vasilescu, M., Terzopoulos, D.: Multilinear independent component analysis. In: CVPR, vol. 1, pp. 547–553 (2005)
8. Papademetris, X., Jackowski, M., Rajeevan, N., Okuda, H., Constable, R., Staib, L.: Bioimage suite: An integrated medical image analysis suite, section of bioimaging sciences, dept. of diagnostic radiology, yale university
9. Papademetris, X., Sinusas, A.J., Dione, D.P., Constable, R.T., Duncan, J.S.: Estimation of 3-d left ventricular deformation from medical images using biomechanical models. IEEE TMI 21(7), 786–800 (2002)
10. Friedman, J., Hastie, T., Tibshirani, R.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer, New York (2001)
11. Cremers, D.: Dynamical statistical shape priors for level set-based tracking. IEEE TPAMI 28(8), 1262–1273 (2006)
12. Shankar, P.M.: A general statistical model for ultrasonic backscattering from tissues. UFFC 47(3), 727–736 (2000)

A Statistical Model of Right Ventricle in Tetralogy of Fallot for Prediction of Remodelling and Therapy Planning

Tommaso Mansi[1], Stanley Durrleman[1], Boris Bernhardt[2], Maxime Sermesant[1], Hervé Delingette[1], Ingmar Voigt[3], Philipp Lurz[4], Andrew M. Taylor[4], Julie Blanc[5], Younes Boudjemline[5], Xavier Pennec[1], and Nicholas Ayache[1]

[1] INRIA-Méditerranée, Asclepios Project, Sophia Antipolis, France
[2] Neuroimaging of Epilepsy Laboratory, McGill University, Montreal Neurological Institute, Quebec, Canada
[3] Siemens AG, CT SE 5 SCR2, Erlangen, Germany & Chair of Pattern Recognition, University of Erlangen-Nuremberg, Erlangen, Germany
[4] UCL Institute of Child Health & Great Ormond Street Hospital for Children, London, United Kingdom
[5] Service de Cardiologie Pédiatrique, Hôpital Necker-Enfants Malades, Paris, France

Abstract. Patients with repaired Tetralogy of Fallot commonly suffer from chronic pulmonary valve regurgitations and extremely dilated right ventricle (RV). To reduce risk factors, new pulmonary valves must be re-implanted. However, establishing the best timing for re-intervention is a clinical challenge because of the large variability in RV shape and in pathology evolution. This study aims at quantifying the regional impacts of growth and regurgitations upon the end-diastolic RV anatomy. The ultimate goal is to determine, among clinical variables, predictors for the shape in order to build a statistical model that predicts RV remodelling. The proposed approach relies on a *forward* model based on currents and LDDMM algorithm to estimate an unbiased template of 18 patients and the deformations towards each individual shape. Cross-sectional multivariate analyses are carried out to assess the effects of body surface area, tricuspid and transpulmonary valve regurgitations upon the RV shape. The statistically significant deformation modes were found clinically relevant. Canonical correlation analysis yielded a generative model that was successfully tested on two new patients.

1 Introduction

Tetralogy of Fallot (ToF) is a severe congenital heart defect that requires surgical repair early in infancy. Yet, pulmonary valves may be damaged by the surgery, causing chronic regurgitations. As a result, the right ventricle (RV) dilates extremely, its shape is altered and the cardiac function is impaired: new valves must be implanted in adulthood to reduce risk factors [1]. Understanding and quantifying RV remodelling in repaired ToF patients is crucial for patient management and therapy planning. However, high variability in pathology course and in RV anatomy makes difficult the decision of optimal timing for re-intervention [1].

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 214–221, 2009.

Contrary to the left ventricle, whose shape and deformations under pathological conditions are well documented, RV anatomy is complex and can vary tremendously among ToF patients. Several studies investigated possible correlations between clinical parameters in ToF [1]. However, few works have quantified the anatomical alterations of the RV and their evolution due to the disease [2,3]. In [2], the authors measure the most striking differences in RV shape with respect to normals, quantifying some features of the complex RV remodelling observed in ToF. However, only one-dimensional indices are considered despite the availability of 3D segmentations. In [3], the authors present a 4D Active Appearance Model of the beating heart to segment RV in MRI. New indices based on the shape modes are proposed to classify patients from normal. Yet, the authors do not correlate their model with clinical features of ToF.

The clinical challenges raised by ToF encourage applying image-based shape analysis techniques to model the RV anatomical alterations due to pathological factors. These techniques generate a representative template of a population of interest and assess how it deforms within this population [4,5,6,7]. Yet, correlating shape with clinical variables require a rigourous framework: Biases may appear if the template is not defined in a consistent way, which may yield drastic differences in the statistical conclusions. Two strategies are available to create the template. The *backward* approach consists in modelling the template as the average of the deformed observations plus some residuals [4,5]. Such a template can be computed efficiently but the model parameters, especially the residuals, are more difficult to identify. The *forward* approach consists in modelling the observations as deformations of the template plus some residuals [6,7]. Computing the template is more complex but model parameters can be faithfully estimated from images and clinical data.

In view of assisting the cardiologists in establishing the best time for reintervention, we aim at statistically predict the RV remodelling in ToF. As a first step, we propose in this work to quantify the regional impacts of growth and regurgitations upon the end-diastolic RV anatomy of a cohort of 18 young ToF patients. The main deformation modes are estimated using the *forward* approach and analysed through cross-sectional multivariate methods. We then derive a generative model of RV remodelling and test it on two new patients.

2 Methods

The right ventricle (RV) of multiple patients is segmented from cine-MRI as described in Sec. 3.1. To analyse this population of shapes, an unbiased template is first built. This template serves as reference atlas to determine the deformations towards each individual shape. Then, Principal Component Analysis (PCA) is applied on these deformations to extract the main deformation modes. The importance of each mode is statistically assessed with respect to child growth and valvar regurgitation severity, yielding a generative model of RV remodelling. Finally, we investigate how this model can predict the evolution of shape with respect to body surface area.

2.1 Unbiased Template of the Right Ventricle in Tetralogy of Fallot

The RV template is created using the *forward* strategy proposed by [7]. This approach is particularly suited for our purposes as 1) it is non-parametric, shapes are represented by currents; 2) model parameters are well-defined and can be estimated from clinical data, thus enabling statistical analyses; 3) template and deformations are computed simultaneously and consistently and 4) new patients can be integrated in the study without re-estimating the template.

The RV surfaces $\mathcal{T}^i$, or shapes, are modelled as the sum of a diffeomorphic deformation ϕ^i of the template $\overline{\mathcal{T}}$ and a residual term ϵ^i standing for the shape features that cannot be represented by the template (topology changes, acquisition artefacts, etc.): $\mathcal{T}^i = \phi^i.\overline{\mathcal{T}} + \epsilon^i$. Currents are used to represent the shapes, the residuals and the deformations in the same common framework. They enable the usual operations (mean, variance...) on shapes as they form a vector space. Intuitively, currents can be seen as the flux of any vector field $\omega \in W$ across the shapes. W is a vector space of infinite dimension generated by a Gaussian kernel $K_W(\mathbf{x},\mathbf{y}) = \exp(-\|\mathbf{x}-\mathbf{y}\|^2/\lambda_W^2)$ that defines an inner product in W (W is a Reproducible Kernel Hilbert Space, RKHS). More precisely, a triangle centred at $\mathbf{x}$ with normal α is represented by the Dirac delta current $\delta_{\mathbf{x}}^{\alpha}$. Therefore, a discrete mesh is encoded by the sum of the currents of its triangles $\mathcal{T}^i = \sum_k \delta_{\mathbf{x}_k^i}^{\alpha_k^i}$. The residuals ϵ^i are modelled as a Gaussian distribution on the α_k^i. The deformation ϕ^i that registers the template $\overline{\mathcal{T}}$ to the current $\mathcal{T}^i$ is estimated using the Large Deformation Diffeomorphic Mappings (LDDMM) framework [8]. ϕ^i is parametrised by a smooth initial vector speed $\mathbf{v}_0^i$, which also belongs to a Gaussian RKHS V with variance λ_V^2. Moreover, this initial speed vector field is completely defined by the *moment* vectors β^i centred at the same point location as the template moments: $\mathbf{v}_0^i(\mathbf{x}) = \sum_k K_V(\mathbf{x}_k,\mathbf{x})\beta_0^i(\mathbf{x}_k)$. Finally, the template $\overline{\mathcal{T}}$ and the deformations ϕ^i towards each patient are estimated by means of an alternate two-step strategy, initialised with the mean current of the population.

2.2 Characterising Deformation Modes of RV Shapes in ToF

In this work we analyse the deformations ϕ^i only as we mainly focus on the regional changes of the RV anatomy due to ToF. Principal Component Analysis (PCA) is performed directly on the moments β_i to extract the main deformation modes. The elements of the covariance matrix Σ are given by $\Sigma_{ij} =< \mathbf{v}_0^i - \overline{v_0}, \mathbf{v}_0^j - \overline{v_0} >_V = \sum_{\mathbf{x}_k,\mathbf{x}_l}(\beta^i(\mathbf{x}_k) - \overline{\beta}(\mathbf{x}_k))K_V(\mathbf{x}_k,\mathbf{x}_l)(\beta^j(\mathbf{x}_l) - \overline{\beta}(\mathbf{x}_l))$, $\mathbf{x}_k$ being the position of the k^{th} Dirac delta current of $\overline{\mathcal{T}}$. Then, the moment vector γ^m of the initial speed vector $\mathbf{u}_0^m$ related to the mth deformation mode is given by $\gamma^m = \overline{\beta} + \sum_i \mathbf{V}^m[i](\beta^i - \overline{\beta})$. In this equation, $\mathbf{V}_m$ is the mth eigenvector of Σ when the eigenvalues are sorted in decreasing order. Finally, the RV shape of each patient i is characterised by the shape vector $\mathbf{s}^i = \{s_m^i\}_{m=1..M}$ computed from the M first deformation modes, $s_m^i =< \mathbf{v}_0^i, \mathbf{u}_0^m >_V = \sum_{\mathbf{x}_k,\mathbf{x}_l} \beta^i(\mathbf{x}_k)K_V(\mathbf{x}_k,\mathbf{x}_l)\gamma^m(\mathbf{x}_l)$.

2.3 Can We Predict the Shape from Clinical Parameters?

First, cross-sectional analysis of the impact of growth on RV shape was performed. Multiple linear regression between the shape vectors $\mathbf{s}^i$ and body surface area (BSA) was carried out to exhibit the effects of BSA on each deformation mode. An optimal set of modes was selected by iteratively removing the modes with lowest significance, until the p-value of the regression overall significance stopped decreasing. Canonical Correlation Analysis (CCA) was then applied to quantify the amount of variation of each mode when BSA varies. Denoting R the overall correlation coefficient between BSA and shape vectors and ρ the correlation vector relating BSA to each deformation mode, the moments μ of the generative deformation Φ are $\mu = R \sum_k \rho[k]\gamma^k$. Deforming the template $\overline{\mathcal{T}}$ with Φ enables quantifying the average RV remodelling observed in our population.

Second, we assessed the impact of tricuspid and transpulmonary regurgitations on each deformation mode. As regurgitations were quantified by ordinal indices and only 18 subjects were available, we choose to perform two independent and component-by-component analyses to maximise statistical power. Rank-based Kruskal-Wallis analysis of variance was applied. If an effect was found for some deformation modes, post-hoc two-sample rank-based Wilconxon tests were used to determine which levels differed.

All the statistical tests were carried out using the shape vectors $\mathbf{s}^i$ (Sec. 2.2). The level of significance was set at $p < 0.1$ and multiple comparisons were corrected using Bonferroni adjustment.

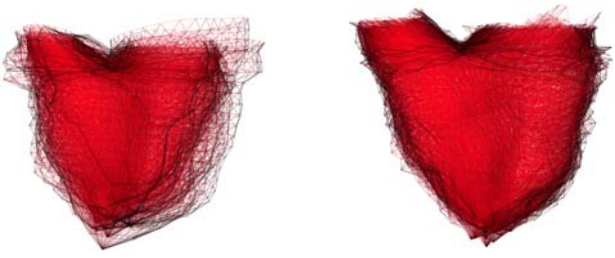
3 Experiments and Results

3.1 Data Collection

Subjects and Image Preparation. We selected 18 young patients (8 males, mean age $\pm$ SD $= 15 \pm 3$) with repaired Tetralogy of Fallot (ToF). Body-surface area (BSA) was reported for each patient (Dubois formula, mean $\pm$ SD $= 1.53 \pm 0.3$). Steady-State Free Precession cine MRI of the heart were acquired with a 1.5T MR scanner (Avanto, Siemens). Images were acquired in the short-axis view covering entirely both ventricles (10-15 slices; isotropic in-plane resolution: 1.1x1.1mm to 1.7x1.7mm; slice thickness: 6-10mm; 25-40 phases). Images were made isotropic through tricubic resampling.

Surface Meshes Preparation. End-diastolic RV endocardium was segmented on the MRI cardiac sequence by fitting an anatomically accurate geometrical model. Its position, orientation and scale in the image was determined using minimal user interaction. Then, local boundaries were estimated by training a probabilistic boosting tree classifier with steerable features [9]. To reduce positioning effects in the shape analysis, the RV meshes were rigidly registered to a representative patient of the dataset by using GMMReg[1] [10]. The results were visually inspected and remaining undesirable rotations were corrected manually (Fig. 1, left panel).

[1] http://code.google.com/p/gmmreg/



Rigid alignment Non-linear registration to the template

Fig. 1. 3D RV meshes of 18 young ToF patients. *Left panel:* The meshes were rigidly registered to a representative patient of the dataset. Observe the extreme variability in shape *(see companion video). Right panel:* The same meshes registered to the template using the non-linear deformations estimated during the template creation.

3.2 Statistical Shape Model of the Right Ventricles

Building the template $\overline{\mathcal{T}}$ required setting two parameters (Sec. 2.1): λ_V, which defines the "stiffness" of the non-linear deformations (the higher is λ_V, the more global are the transformations); and λ_W, which characterises the resolution of the currents representation (low λ_W values enable analysing subtle shape features). As we were mainly interested in the regional ToF alterations (dilation, valve enlargement, regional bulging), these parameters were set to $\lambda_W = \lambda_V = 15$mm, about the diameter of the RV outflow tract. Lower values would have been inappropriate as the image slice thickness was approximately 10mm.

One iteration of the alternate minimisation was needed to reach convergence. Yet, the resulting template $\overline{\mathcal{T}}$ was well centred (mean over standard deviation of the deformations was 0.8). The first 10 deformation modes were selected, representing more than 90% of the total energy (Fig. 2).

Interestingly, the age of the closest patient to the template was 17 while his BSA 1.76. Both indices were close to the observed mean, suggesting that in our population, the mean shape was consistent with the mean BSA and age. Furthermore, this patient only suffers from trace valvar regurgitations, which is not surprising as no evident pathological bulging were visible in the template.

3.3 Statistical Model of RV Remodelling in ToF Patients

Patient growth was quantified by body surface area (BSA) index (correlation with age in the data set: $R^2 > 0.5$, $p < 0.001$). Table 1 reports the regression coefficients a_l of the multiple linear regression between BSA and shape vectors $\mathbf{s}$, $\text{BSA} = a_0 + \sum_{l=1}^{10} a_l \mathbf{s}[l]$, the related t-values and the overall model significance. The sign of the a_l relates to the direction of the deformation modes (negative a_l meaning backwards deformations). Model reduction discarded all the non-significant modes (Table 1). The remaining modes were found clinically pertinent by an expert after visual inspection (Fig. 2). Mode 1 clearly represented the overall RV dilation. Mode 2 seemed to model the dilation of the tricuspid annulus and of the inflow tract. Mode 3, 6, 7 and 9 exhibited a dilation of a specific RV

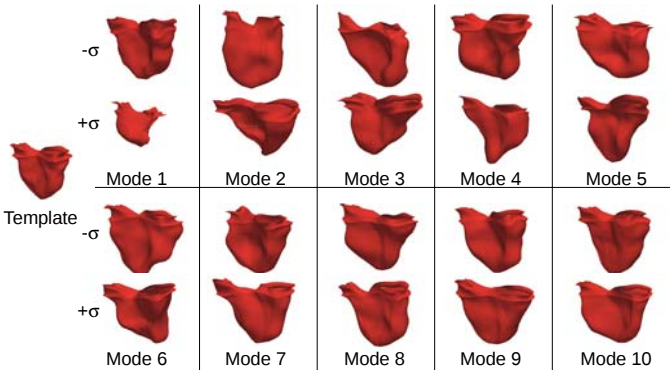


Fig. 2. 10 first deformation modes extracted by PCA on a population of 18 patients suffering from repaired Tetralogy of Fallot

Table 1. Linear regression coefficients a_l between shape modes and BSA. In bold the significant coefficients ($p < 0.1$). After model reduction (second array), coefficients stay unchanged, confirming the stability of the statistical tests.

	Significance	$\mathbf{a_1}$	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}
Coef. $\times 10^{-5}$	$R^2 = 0.84$,	**-2.9**	**6.4**	**-7.6**	4.6	-1.0	**-11.1**	**-11.9**	7.0	**-20.1**	-15.4
t-values	$p = 0.04$	-3.28	2.64	-2.04	1.13	-0.19	-1.93	-1.92	0.84	-2.15	-1.43
Coef. $\times 10^{-5}$	$R^2 = 0.75$,	**-2.9**	**6.4**	**-7.6**			**-11.1**	**-11.9**		**-20.0**	
t-values	$p = 0.006$	-3.27	2.63	-2.03			-1.92	-1.92		-2.14	

region: apex (mode 3), basal area under the tricuspid valve (mode 6), apical area of the outflow tract (mode 7) and outflow tract (mode 9), reflecting possible direct impact of regurgitations on the neighbouring tissues.

Canonical Correlation Analysis (CCA) provided a generative model of the RV remodelling observed in our population. Overall correlation coefficient with BSA was $R = 0.87$, suggesting a strong correlation between these deformation modes and growth. The correlation vector of the deformation modes was $\rho = \{-0.56, 0.45, -0.35, -0.33, -0.33, -0.37\}$. When BSA increases by 0.86, each deformation mode m varies by its related coefficient $\rho[m]$. The model was found clinically realistic by an expert (Fig. 3). As BSA increased, RV volume increased, RV free-wall and valves dilated, and septum was more concave. Indeed, dilation of the valves reduces the remaining pulmonary obstructions, thus decreasing the RV pressure. As a result, left-ventricle pushes the septum towards the right ventricle, making it more concave. However, as regurgitations are still present, the RV still dilates by pushing the RV free wall outwards.

3.4 Quantifying the Impact of Valvar Regurgitations on RV Shape

Colour Doppler ultrasound (sweep speeds: 50-100 mm/s) was used to quantify tricuspid (TriReg) and transpulmonary valve (TPVReg) regurgitations. To assess the effects of TPVReg, patients were grouped into two different groups:

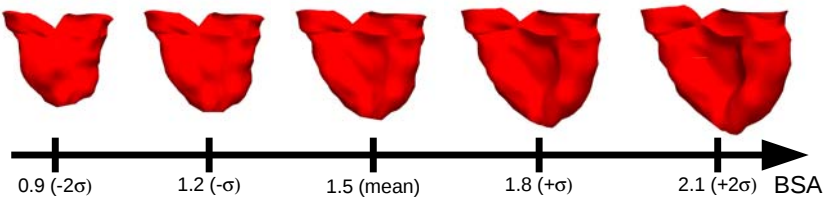


Fig. 3. Mean RV remodelling observed in our population when body surface area (BSA, in m^2) increases. RV dimensions globally increase while valves dilate. Simultaneously, RV free wall becomes rounder and septum more concave *(see companion videos)*.

trace TPVReg and severe TPVReg. Kruskal-Wallis analysis revealed a significant effect on deformation mode 2 ($p < 0.1$), which was confirmed by visual inspection: this mode exhibited an elongation of the RV outflow tract (Fig. 2).

Evaluation of TriReg classified the patients into 3 groups: none, trace or mild tricuspid regurgitations. Kruskal-Wallis analysis showed a significant impact of TriReg on three deformation modes: 3 ($p < 0.05$), 6 ($p < 0.1$) and 8 ($p < 0.1$). However, pair-wise Wilconxon tests showed that only mode 3 had two significant different levels (trace TriReg versus mild TriReg, $p < 0.1$). Visually, deformation mode 3 exhibited a deformation of the tricuspid annulus, from circular to triangular-shape, and a dilation of the RV inflow tract.

Interestingly, two deformation modes involved in the statistical model of RV growth were also related to the regurgitations. This may suggest possible cross-effects between growth and regurgitations on these specific shape variations.

3.5 Validating the Generalisation of the Statistical Models

Generalising a statistical model of RV remodelling is crucial for patient management and therapy planning. We thus tested the robustness of our model on two new patients with matched age (13 and 16). The template was registered to the patients and the related shape vectors $\mathbf{s}$ were computed. BSA were calculated from the optimal linear model estimated in Sec. 3.3. Results successfully compared with measured values (patient 1: estimated BSA: 1.61, measured BSA: 1.49; patient 2: estimated BSA: 1.29, measured BSA: 1.16). This suggests that the deformation modes involved in the RV remodelling model could be generalised, constituting potential quantitative parameters of remodelling in ToF.

4 Discussion and Future Works

In this study we investigated the impact of growth and valvar regurgitations upon the end-diastolic RV anatomy of patients with repaired ToF. End-diastolic time point was chosen as it is the time when the effects of the pathology are the most evident [1,2]. Multivariate statistical analyses provided a generative model of the observed RV remodelling. This model and the significant deformation modes were found clinically pertinent as they exhibited remarkably realistic changes in RV anatomy. To design the model, the deformation modes and their

directions were statistically determined to limit the effects of PCA rotatability. Furthermore, the effects of regurgitation severity were analysed on a component-by-component basis to preserve the statistical power of the tests due to the ordinal nature of the data. The groups were not sufficiently populated to apply more comprehensive statistics. Incorporating more patients is now required to confirm these findings and avoid possible over-interpretations. Various types of RV remodelling could be identified (aneurysmal, with stiff myocardium, etc.), which may constitute new criteria for valve replacement decision. Future works also include analysing the 4D cardiac motion. To the best of our knowledge, this study constitutes a first attempt at correlating 3D shape parameters to clinical measurements in ToF. These analyses may yield quantitative image-based predictors about RV anatomy and remodelling in ToF.

Acknowledgments. This work has been partly funded by the European Commission, IST-2004-027749 Health-e-Child Integrated Project. The authors warmly thank Pr. G. Pongiglione, Istituto G. Gaslini, Genova, for his helpful comments about the models.

References

1. Geva, T.: Indications and timing of pulmonary valve replacement after tetralogy of Fallot repair. In: Seminars in Thoracic and Cardiovascular Surgery: Pediatric Cardiac Surgery Annual, vol. 9, pp. 11–22. Elsevier, Amsterdam (2006)
2. Sheehan, F., Ge, S., Vick III, G., Urnes, K., Kerwin, W., Bolson, E., Chung, T., Kovalchin, J., Sahn, D., Jerosch-Herold, M., et al.: Three-Dimensional Shape Analysis of Right Ventricular Remodeling in Repaired Tetralogy of Fallot. The American Journal of Cardiology 101(1), 107 (2008)
3. Zhang, H., Walker, N., Mitchell, S., Thomas, M., Wahle, A., Scholz, T., Sonka, M.: Analysis of four-dimensional cardiac ventricular magnetic resonance images using statistical models of ventricular shape and cardiac motion. In: Proc. SPIE 2006, vol. 6143, p. 614307 (2006)
4. Guimond, A., Meunier, J., Thirion, J.P.: Average brain models: A convergence study. Computer Vision and Image Understanding 77(2), 192–210 (2000)
5. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)
6. Allassonniere, S., Amit, Y., Trouve, A.: Towards a coherent statistical framework for dense deformable template estimation. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 69(1), 3–29 (2007)
7. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: A forward model to build unbiased atlases from curves and surfaces. In: Pennec, X., Joshi, S. (eds.) Proc. of the International Workshop on the Mathematical Foundations of Computational Anatomy, MFCA 2008 (2008)
8. Vaillant, M., Glaunès, J.: Surface matching via currents. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 381–392. Springer, Heidelberg (2005)
9. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Fast automatic heart chamber segmentation from 3D CT data using marginal space learning and steerable features. In: Proc. ICCV 2007, pp. 1–8 (2007)
10. Jian, B., Vemuri, B.: A robust algorithm for point set registration using mixture of Gaussians. In: Proc. ICCV 2005, vol. 2, pp. 1246–1251 (2005)

Bayesian Maximal Paths for Coronary Artery Segmentation from 3D CT Angiograms

David Lesage[1,2], Elsa D. Angelini[2], Isabelle Bloch[2], and Gareth Funka-Lea[1]

[1] Siemens Corporate Research, Imaging and Visualization dept., Princeton NJ, USA
[2] Institut Telecom, Telecom ParisTech, CNRS LTCI, Paris, France

Abstract. We propose a recursive Bayesian model for the delineation of coronary arteries from 3D CT angiograms (cardiac CTA) and discuss the use of discrete minimal path techniques as an efficient optimization scheme for the propagation of model realizations on a discrete graph. Design issues such as the definition of a suitable accumulative metric are analyzed in the context of our probabilistic formulation.

Our approach jointly optimizes the vascular centerline and associated radius on a 4D space+scale graph. It employs a simple heuristic scheme to dynamically limit scale-space exploration for increased computational performance. It incorporates prior knowledge on radius variations and derives the local data likelihood from a multiscale, oriented gradient flux-based feature. From minimal cost path techniques, it inherits practical properties such as computational efficiency and workflow versatility. We quantitatively evaluated a two-point interactive implementation on a large and varied cardiac CTA database. Additionally, results from the Rotterdam Coronary Artery Algorithm Evaluation Framework are provided for comparison with existing techniques. The scores obtained are excellent (97.5% average overlap with ground truth delineated by experts) and demonstrate the high potential of the method in terms of robustness to anomalies and poor image quality.

1 Introduction

In many biomedical applications, the segmentation of vascular structures is an important step towards diagnosis, treatment and surgery planning. Modern 3D angiographic modalities produce increasingly large and detailed acquisitions. This exacerbates the need for automated or semi-automated methods to reduce the burden of manual delineation while increasing repeatability.

Among the rich literature on vascular segmentation [1], *minimal path* methods are particularly popular. They classically extract a vessel as a path between two points on a regular lattice. The cost of a path is defined cumulatively by a vessel-dedicated metric and optimized over the lattice by graph-based methods (*e.g.*, Dijkstra-based and fast-marching algorithms for L_1 and L_2 metrics, respectively). Recent works [2,3] have shown that the lattice is not necessarily limited to the spatial positions of the vessel centerline and can be augmented with dimensions such as the vessel radius. Key properties of minimal paths are

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 222–229, 2009.

their guarantee of global optimality, their computational efficiency and the control over boundary conditions. One can always find a path between two given points or even search exhaustively, *i.e.*, extract the paths from a single seed to *all* the other points. Most drawbacks, such as so-called "shortcut" issues, come from the practical difficulty of designing well-behaved cumulative metrics.

Another particularly active class of methods is the family of multi-hypotheses tracking methods [4,5,6,7]. They increase the robustness of the local tracking process by evolving several hypotheses in parallel. They differ on how hypotheses are selected and evolved. Bayesian formulations such as particle filters [5] and related schemes [6,7] have demonstrated their robustness and design versatility, in particular in the integration of prior model-based knowledge. Unlike minimal path techniques, they do not offer control over the exhaustiveness of the search, but are designed to explore large search spaces in a sparse manner, focusing on most promising areas. They are generally not dependent on the discretization of the search space but can be computationally expensive.

In this paper, we introduce a recursive Bayesian model related to those underlying multi-hypothesis probabilistic methods [5,6,7]. We discuss the use of minimal path techniques as an efficient optimization scheme propagating model realizations as paths on a discrete graph. We detail key design aspects such as the definition of the data likelihood and the derivation of a cumulative metric from the Bayesian formulation. Our approach was applied to the particularly challenging task of segmenting coronary arteries from 3D cardiac CTA. Qualitative and quantitative evaluation is given on clinical data.

2 Geometric and Bayesian Models

A vascular segment is modeled as a discrete series of states $X_{t\in[0:T]} = \{(p_t, r_t)\}$, composed of centerline points p_t and associated radiuses r_t (Fig. 1). Cross-sections are assumed to be circular, a reasonable approximation for thin vessels such as coronaries. Their local orientation is approximated by $d_t = \frac{p_t - p_{t-1}}{|p_t - p_{t-1}|}$.

We are interested in $X^* = \arg\max_X P(X)P(Y|X)$, the *maximum a posteriori* model realization given image observations $Y = \{Y^j\}$. An observation Y^j is obtained as the response of a multiscale oriented feature, computed at location p^j, radius r^j and direction d^j (Sec. 4). The likelihood $P(Y^j|X = X^i_{[0:t]})$ depends on whether an observation Y^j is associated to the model realization $X^i_{[0:t]}$ or not. A particularity of our model is to distinguish two distributions, $P_v(Y^j)$ and

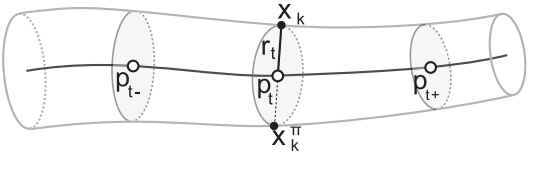


Fig. 1. Geometric model and parameterization

$P_{bg}(Y^j)$ for responses in vessels and in the background, respectively (similarly to [8] for road tracking). Assuming conditionally independent observations given a model realization, the joint likelihood of the observations is:

$$P(Y|X^i_{[0:t]}) = \prod_{Y^j \in X^i_{[0:t]}} P_v(Y^j) \prod_{Y^j \notin X^i_{[0:t]}} P_{bg}(Y^j) = \prod_{Y^j \in X^i_{[0:t]}} \frac{P_v(Y^j)}{P_{bg}(Y^j)} \prod_{Y^j, \forall j} P_{bg}(Y^j)$$

where the (abusive) notation $Y^j \in X^i_{[0:t]}$ indicates that Y^j is associated to a certain state of $X^i_{[0:t]}$. We note $Y^j = Y^i_t$ if Y^j is associated with the particular state X^i_t. Assuming a 1^{st} order Markovian prior and omitting the terms independent of the model realization, the recursive update of the *a posteriori* probability of a model realization (also referred to as an *hypothesis*) is given by Bayes' rule:

$$P(X^i_{[0:t]}|Y^i_{[0:t]}) \propto P(X^i_{[0:t-1]}|Y^i_{[0:t-1]}) P(X^i_t|X^i_{t-1}) \frac{P_v(Y^i_t)}{P_{bg}(Y^i_t)} \quad (1)$$

This Bayesian model is similar to those used in recent works on vascular probabilistic tracking [5,6,7]. The main difference is the integration of background information in the likelihood expression, similarly to [8].

3 Graph Optimization for Hypothesis Propagation

The theoretical maximum *a posteriori* problem is generally intractable as the search space grows exponentially with the length T of the model. In practice, probabilistic multi-hypothesis methods adopt *selection* schemes to limit the number of hypotheses, *e.g.*, stochastic resampling [5,7] or deterministic pruning [6].

As an alternative, we rely on the discretization of the search space, allowing the use of an efficient and robust minimal path-like optimization scheme. Spatial positions $\{p^i_t\}$ are discretized as the regular image grid and we use R different radius values (see Sec. 4). Our approach thus explores a 4D space+radius graph, where each node corresponds to a possible state $X^k = (p^i, r^j)$. To value the edges of the graph, we define our additive cost metric by noting that maximizing (1) is equivalent to minimizing its negative logarithm:

$$C(X^i_{[0:t]}) = C(X^i_{[0:t-1]}) - \log(P(X^i_t|X^i_{t-1})) - \log(P_v(Y^i_t)) + \log(P_{bg}(Y^i_t)) - M \quad (2)$$

where $C(X^i_{[0:t]}) = -\log(P(X^i_{[0:t]}|Y^i_{[0:t]}))$ and M is a constant ensuring that $C(X^i_{[0:t]})$ remains positive or null[1]. This additive cost metric is directly suitable for Dijkstra-like optimization. Our algorithm sorts the hypotheses in a min-heap structure. When the heap root $X^i_t = (p^i_t, r^i_t)$ is popped, we consider its neighbor states, defined by the product set of the 26-neighboring positions $\{p^i_{t+1}\}$ of p^i_t and all possible scales. The cost of each neighbor is updated according to (2) (details in Sec. 4). Propagation continues until the heap is empty (exploration

[1] We set $M = log(P^{min}_{bg})$, with $P^{min}_{bg} = 10^{-6}$ in our implementation.

of the entire grid) or when a given end position is popped (two-seed workflow). The result path (points+radiuses) is simply backtracked from the end state.

Exploring the full 4D graph can be very costly in time and memory. To alleviate that cost, we propose to propagate only the $H < R$ best hypotheses (of different radiuses) first reaching each spatial position. H controls the local, dynamic radius selection scheme limiting scale-space exploration. For $H = R$, the optimization would be globally optimal. With $H = 1$, only the locally best radius is propagated. Using $H > 1$ increases robustness to scale-related ambiguities.

Our approach can be viewed as a generalization of [9]. We also derive edge costs from an oriented, multiscale medialness feature, but our method uses the feature response indirectly to value the likelihood terms, incorporates radiuses in the optimization and controls the coherence of their variations through the prior term (Sec. 4). From its space+scale minimal path approach, our method is closely related to [3]. Where our Bayesian model yields an edge-based L_1 metric, authors in [3] prefer L_2 optimization, arguably less sensitive to discretization. We believe however that our metric is less parameter-dependent than the node-based potentials proposed in [3]. More importantly, our algorithm does not explore the 4D graph entirely and rather focuses on most promising radiuses, yielding considerable memory and time gains in practice. For $H < R$, our optimization scheme is heuristic and does not guarantee the extraction of the global maximum, but it is computationally efficient, based on a sound theoretical framework and particularly robust, as confirmed by the validation results from Sec. 5.

4 Implementation Details

Our implementation relies on an interactive workflow with user-provided start and end points. The heap is initialized with the set of seed states $X_0^i = (p_0, r^i)$, *i.e.* all possible radiuses r^i for the start position p_0. To cover the typical radius range of coronary arteries, we use R=16 radius values from $r_{\min}$=0.3 to $r_{\max}$=5.1mm with a fixed step of 0.3mm.

The prior term $P(X_t^i|X_{t-1}^i)$ in (2) constrains radius variations:

$$P(X_t^i|X_{t-1}^i) = P(r_t^i|r_{t-1}^i) = \begin{cases} (\sigma_-\sqrt{2\pi})\mathcal{N}(r_t^i|r_{t-1}^i, \sigma_-^2) & \text{if } r_t^i \leq r_{t-1}^i, \\ (\sigma_+\sqrt{2\pi})\mathcal{N}(r_t^i|r_{t-1}^i, \sigma_+^2) & \text{if } r_t^i > r_{t-1}^i. \end{cases} \tag{3}$$

where $\mathcal{N}(.|\mu, \sigma^2)$ is the Normal distribution of mean μ and variance σ^2. We use an asymmetric formulation to penalize widening models more than shrinking models. We experimentally set σ_-=0.4mm and σ_+=0.2mm. These distributions are kept unnormalized, as discussed at the end of the section.

Observations Y_t^i are the responses of a multiscale oriented medialness feature:

$$Y_t^i = \texttt{MFlux}(p_t^i, r_t^i, d_t^i) = \frac{2}{N}\sum_{k=1}^{\frac{N}{2}} \min(\langle \nabla I(x_k), u_k \rangle, \langle \nabla I(x_k^\pi), u_k^\pi \rangle) \tag{4}$$

At cross-sectional contour points x_k, `MFlux` [10] accumulates the projections of the image gradient vectors $\nabla I(x_k)$ on inward radial directions $u_k = p_t^i - x_k$. The

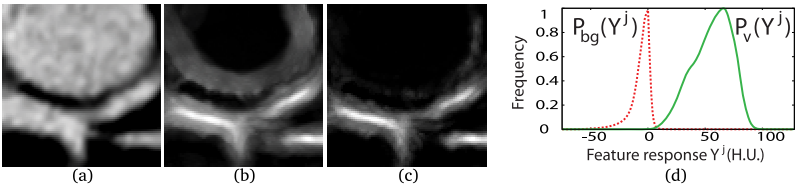


Fig. 2. Flux-based feature and likelihood distributions. (a) Detail of a dataset with arteries running close to a heart chamber. (b) Response of the local cross-sectional gradient flux (maximum over scales and directions for each voxel). (c) Response of `MFlux`. (d) Examples of P_v and P_{bg} likelihood distributions.

sum of these projections corresponds to a measure of the local inward gradient flux through the cross-section, which is maximal when the model cross-section is well aligned with a vessel in the image. Instead of taking the sum, `MFlux` retains the minimum contribution between diametrically opposed points x_k and x_k^π (see Fig. 1). This non-linear combination dramatically reduces false positive step-edge responses caused by asymmetric flux contributions at locations such as the surface of heart chambers (Fig. 2). `MFlux` enjoys a high discriminative power over scales and positions while remaining computationally efficient. we use N=8 cross-section points and precompute gradient vectors using Gaussian derivatives with σ_g=0.3mm. Please refer to [10] for more in-depth discussion.

Observations are used to value the likelihood terms $P_v(Y_t^i)$ and $P_{bg}(Y_t^i)$ in (2). The vessel likelihood P_v was learnt as the response histogram on 10 datasets with ground truth segmentation (not included in the validation set of Sec. 5). The background likelihood P_{bg} is dataset-dependent and is estimated through feature responses at 10^6 randomly sampled parameters (p, r, d). It thus encodes general information about the reliability of the data at hand, as it will vary as a function of the image quality and noise. P_v and P_{bg} are generally well separated thanks to the discriminative power of `MFlux`, as depicted in Fig. 2 (d).

We finally discuss the normalization of the distributions in (2), which can be considered as an implicit parameterization of our metric. Although (2) is defined up to an additive constant, the scaling of the distributions greatly influences how model realizations are compared. In particular, our algorithm deals with hypotheses of different lengths competing for the same position. Instead of classically normalizing each distribution so that it sums to unity, we scale them so that their respective maximum is always 1. Consequently, an ideal model realization will have a cumulative cost (log-probability) of 0 regardless of its length. By limiting length-related penalization to its minimum, we effectively reduce the risk of "shortcuts", a classical issue of minimal path techniques.

5 Experiments and Validation

We first discuss some of the distinctive components of our approach. In general, the background likelihood $P_{bg}(Y_t^i)$ has a limited impact on the extracted result, but it markedly decreases the extent of exploration needed to reach the end state

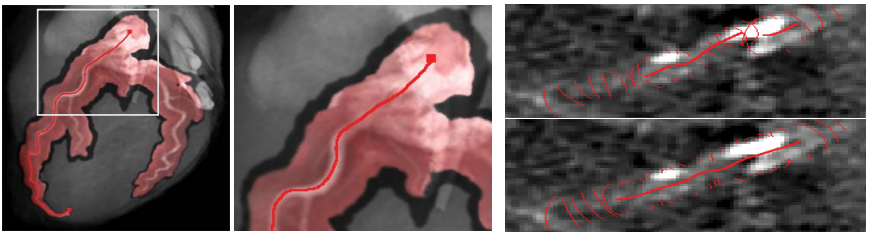


Fig. 3. Effect of the background likelihood and scale prior terms. Left and middle: search space explored to extract the red vessel. Highlighted in red: *with* background likelihood P_{bg} (1.1×10^6 voxels visited). In dark: *without* P_{bg} (1.95×10^6 voxels). Right: difficult case extracted *without* scale prior (top) and *with* scale prior (bottom).

Table 1. Summary of quantitative results (see text for details)

(a) Internal validation

Measure	H=1	H=2	H=4
OV (avg.)	97.8%	99.1%	**99.3%**
AD (avg.)	0.375mm	0.365mm	**0.365**mm
AR (avg.)	0.195mm	0.192mm	**0.191**mm

(b) Results from the Rotterdam challenge [11]

Measure	% / mm			score		
	min.	max.	avg.	min.	max.	avg.
OV	76.0%	100.0%	**97.5%**	39.7	100.0	**81.5**
AD	0.20mm	3.78mm	**0.37**mm	22.4	55.9	**36.3**

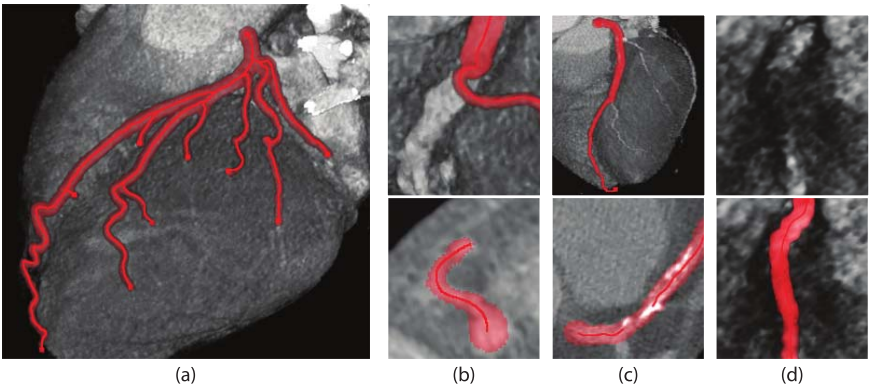


Fig. 4. Result samples (extracted centerlines and mask from local radius estimation). (a) Full arterial tree extracted with one ostia seed and 9 distal seeds (in 57 sec. with $H = 4$). (b) 3D and multi-planar reformation (MPR) views of a curvy, small secondary vessel branching off the right coronary artery. (c) 3D and MPR views of a calcified left coronary artery. (d) Low image quality and occluded right coronary artery.

(up to ∼50%, see Fig. 3). By penalizing hypotheses in non-vascular areas, it yields significant speed gains. The scale prior $P(r_t^i | r_{t-1}^i)$ improves the robustness of the extraction in presence of local anomalies and/or low image quality (see Fig. 3, right) while still allowing smooth scale adaptations (Fig. 4 (b)).

Our algorithm was first validated on 51 CTA datasets of varying image quality and pathologies (Fig. 4). This database counts 858 coronary branches manually delineated by experts, including radius estimation. Table 1(a) shows quantitative

results for different values of H. OV is the average overlap with ground truth, *i.e.*, the portion of ground truth and extracted points coinciding within radius distance. The version with H=1 already obtains a high overlap score of 97.8%, improved up to 99.3% for H=4. These results demonstrate the high robustness of our approach and the relevance of keeping several hypotheses per spatial location. Increasing H over 4 did not bring significant improvement in our tests. AD, the average distance of the result centerline to the ground truth, is slightly larger than the typical intra-slice data resolution (0.33mm per voxel). This accuracy result is coherent with the discrete nature of our algorithm. The average radius estimation error AR of 0.191mm is satisfyingly subvoxelic.

We additionally submitted results (for $H = 4$) to the publicly open Rotterdam Coronary Artery Algorithm Evaluation Framework [11]. In Table 1(b), we report the average overlap OV and average distance AD results for 24 testing datasets (96 vessels). Other statistics are left out due to space restrictions [12]. Scoring in $[0, 100]$ is defined as follows: 50 for a result of the order of the inter-observer variability, 100 for a perfect result. The good performance of our method is confirmed, with 97.5% average overlap, exceeding inter-observer overlap (score of 81.5). The average distance AD is slightly larger than inter-observer variability (score of $36.3 < 50$). Our primary focus being robustness, we consider this accuracy level to be satisfactory for initial delineation before subsequent refinement. This evaluation also allows the direct comparison with other existing methods. For instance, our algorithm brings a noticeable robustness improvement (97.5% versus 91.9% OV) over the method from [13] which implements a classical two-seed minimal path technique based on image intensity and a Hessian-based vesselness feature. To date, the only publicly ranked method to outperform our approach in terms of robustness is [14], an adaptation of [4] supplemented with minimal paths, which obtained $\mathtt{OV} = 98.5\%$. It is worth noticing that this method required more interaction, with the use of intermediate seed points, where we strictly limited ourselves to the provided start and end points.

Finally, we emphasize the computational efficiency of our approach. The effort required to extract a vessel depends on several factors, such as the image quality, the length and complexity of the target path and overall vascular network. In order to extract the longest vessel branch, the algorithm will basically explore the entire network (see Fig. 3). Consequently, the entire arterial tree can be extracted for the same computational effort as its longest branch, by specifying one seed at the ostium and one distal point per branch (see Fig. 4). With our C++ implementation, the average vessel branch is extracted in about one minute for $H = 4$ (less than 15 sec. for $H = 1$) on a 2.16GHz Core Duo CPU.

6 Conclusion and Perspectives

In this paper, we have presented a new algorithm for the segmentation of coronary arteries from 3D cardiac CTA data. Our approach relies on a recursive Bayesian model, from which it inherits its robustness, and is optimized by minimal path-like techniques, from which it inherits its computational efficiency and

workflow flexibility. In particular, our algorithm improves on classical minimal path techniques by providing a sound theoretical framework for the definition of the cumulative cost metric, whose different components were carefully studied. The high practical robustness of our technique was demonstrated through an extensive quantitative validation on clinical data.

The very promising results obtained by our proof-of-concept implementation open several high-potential perspectives. One lead is to exploit the reliability of the approach to improve overall clinical workflow for coronary disease assessment through fast, intuitive and reliable interactive tools to extend and correct centerlines for the most difficult cases. Another lead is further automation for full tree extraction. The main difficulty in this case is the design of robust stopping criteria for the propagation. We finally highlight the generality of the approach. Within the same framework, we are currently evaluating other vascular applications by adapting components such as the prior and likelihood distributions.

References

1. Kirbas, C., Quek, F.: A review of vessel extraction techniques and algorithms. ACM Computing Surveys 36(2), 81–121 (2004)
2. Wink, O., Niessen, W., Viergever, M.: Multiscale vessel tracking. IEEE Trans. Med. Imaging 23(1), 130–133 (2004)
3. Li, H., Yezzi, A.: Vessels as 4-D curves: Global minimal 4-D paths to extract 3-D tubular surfaces and centerlines. IEEE Trans. Med. Imaging 26, 1213–1223 (2007)
4. Friman, O., Hindennach, M., Peitgen, H.O.: Template-based multiple hypotheses tracking of small vessels. In: Proc. Int. Symp. Biom. Imaging, pp. 1047–1050 (2008)
5. Florin, C., Paragios, N., Williams, J.: Particle filters, a quasi-Monte Carlo solution for segmentation of coronaries. Med. Image Comput. Assist. Interv., 246–253 (2005)
6. Schaap, M., Smal, I., Metz, C., van Walsum, T., Niessen, W.: Bayesian tracking of elongated structures in 3D images. In: Inf. Process. Med. Imaging, pp. 74–85 (2007)
7. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: Medial-based bayesian tracking for vascular segmentation: Application to coronary arteries in 3D CT angiography. In: Proc. Int. Symp. Biom. Imaging, pp. 268–271 (2008)
8. Geman, D., Jedynak, B.: An active testing model for tracking roads in satellite images. IEEE Trans. Pattern Anal. Mach. Intell. 18(1), 1–14 (1996)
9. Gulsun, M.A., Tek, H.: Robust vessel tree modeling. In: Med. Image Comput. Assist. Interv., pp. 602–611 (2008)
10. Lesage, D., Angelini, E.D., Bloch, I., Funka-Lea, G.: Design and study of flux-based features for 3D vascular tracking. In: Proc. Int. Symp. Biom. Imaging (2009)
11. Rotterdam Coronary Artery Algorithm Evaluation Framework, http://coronary.bigr.nl/
12. Metz, C., Schaap, M., van Walsum, T., van Der Giessen, A., Weustink, A., Mollet, N., Krestin, G., Niessen, W.: 3D segmentation in the clinic: A grand challenge II - coronary artery tracking. The MIDAS Journal (2008)
13. Metz, C., Schaap, M., van Walsum, T., Niessen, W.: Two point minimum cost path approach for CTA coronary centerline extraction. The MIDAS Journal (2008)
14. Friman, O., Kühnel, C., Peitgen, H.: Coronary artery centerline extraction using multiple hypothesis tracking and minimal paths. The MIDAS Journal (2008)

Hierarchical Normalized Cuts: Unsupervised Segmentation of Vascular Biomarkers from Ovarian Cancer Tissue Microarrays

Andrew Janowczyk[1], Sharat Chandran[1], Rajendra Singh[2], Dimitra Sasaroli[3], George Coukos[3], Michael D. Feldman[4], and Anant Madabhushi[5,⋆]

[1] Dept of Computer Science & Engineering, Indian Institute of Technology Bombay
{andrew,sharat}@cse.iitb.ac.in
[2] Quest Diagnostics, Inc, USA
[3] Ovarian Cancer Research Center, University of Pennsylvania, USA
[4] Dept of Pathology and Lab Medicine, University of Pennsylvania, USA
[5] Dept of Biomedical Engineering, Rutgers University, USA
anantm@rci.rutgers.edu

Abstract. Research has shown that tumor vascular markers (TVMs) may serve as potential OCa biomarkers for prognosis prediction. One such TVM is ESM-1, which can be visualized by staining ovarian Tissue Microarrays (TMA) with an antibody to ESM-1. The ability to quickly and quantitatively estimate vascular stained regions may yield an image based metric linked to disease survival and outcome. Automated segmentation of the vascular stained regions on the TMAs, however, is hindered by the presence of spuriously stained false positive regions. In this paper, we present a general, robust and efficient unsupervised segmentation algorithm, termed Hierarchical Normalized Cuts (HNCut), and show its application in precisely quantifying the presence and extent of a TVM on OCa TMAs. The strength of HNCut is in the use of a hierarchically represented data structure that bridges the mean shift (MS) and the normalized cuts (NCut) algorithms. This allows HNCut to efficiently traverse a pyramid of the input image at various color resolutions, efficiently and accurately segmenting the object class of interest (in this case ESM-1 vascular stained regions) by simply annotating half a dozen pixels belonging to the target class. Quantitative and qualitative analysis of our results, using 100 pathologist annotated samples across multiple studies, prove the superiority of our method (sensitivity 81%, Positive predictive value (PPV), 80%) versus a popular supervised learning technique, Probabilistic Boosting Trees (sensitivity, PPV of 76% and 66%).

⋆ This work was supported via grants from the New Jersey Commission on Cancer Research, the National Cancer Institute (R21CA127186-01, R03CA128081-01), Wallace H. Coulter Foundation (PI: Anant Madabhushi) and Ovarian Cancer SPORE Grant P50 CA083638 (PI: George Coukos).

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 230–238, 2009.

1 Introduction

It is estimated[1] that 21,650 women will be diagnosed with and 15,520 women will die of cancer of the ovary (OCa) in 2008. The 5-year survival rates of these women are highly correlated to the early detection of OCa. Recent work [2] suggests that specific tumor vascular biomarkers (TVMs) may be identifiable on OCa TMAs that could have prognostic significance, helping to not only predict this survival rate but also help determine a more specific course of treatment. It has also been suggested that genes expressed uniquely by the vasculature of tumors may provide important therapeutic targets. Biomarkers are typically discovered by staining explicitly for TVMs of interest on OCa TMAs. ESM-1 is one such TVM of interest in OCa, which can be visualized by staining the anti-body to ESM-1 in TMAs with Diaminobenzidine. Precise quantification of the extent and intensity of the stain could serve as a prognostic metric reflecting risk of disease recurrence and patient survival. However, it is currently infeasible in terms of both time and effort for an expert pathologist to perform this segmentation manually.

One of the issues to consider in the choice of a segmentation method is that it should accurately and reliably distinguish true positive (TP) stains from other structures in the tissue that pick up the D-benzidine stain, but are not part of the vasculature. This produces a dataset that suffers greatly when operated on by algorithms that are highly dependent upon initial conditions, such as k-means. Since there is a significant amount of inter-sample and inter-study variability, a more sophisticated algorithm must be employed. In addition, the need to process several hundred cylinders quickly, with high accuracy, is another major requirement. Probabilistic boosting trees (PBT) [3] are popularly employed for classification and segmentation because they have the attractive property that the posterior probability can be used as a threshold to balance between sensitivity and specificity. Additionally, the method is relatively fast. Unfortunately, one of the challenges in constructing a supervised classifier (such as PBTs) is the difficulty in obtaining ground truth segmentations for classifier training for the object, or region, of interest from experts. Creating enough acceptable training samples is both costly and time consuming.

The mean shift (MS) algorithm was originally presented in [4] and revised in [5] as an unsupervised technique aimed at mode discovery for use in place of k-means. MS attempts to overcome its predecessors faults by searching for the mean data point within a certain bandwidth (σ) setting. The improved Fast Gauss Transform [6] implementation of MS allows computation times for large images to become reasonable for use in clinical settings.

The Normalized Cuts (NCut) algorithm is among the final mature descendants of a series of graph cutting techniques ranging from max cut to min cut [7]. It is a popular scheme in spite of its main drawback: the large number of calculations needed for determining the affinity matrix and the time consuming eigenvalue computation. In large sized images the computation and overhead of these border on the infeasible [8], leading to a significant amount of research in avoiding

[1] Cancer Facts and Figures, American Cancer Society (ACS), Atlanta, Georgia, 2008.

their direct calculations [9]. A standard approach in the NCut method is to make the assumption that pixels in the same cluster are spatially located near each other, resulting in a sparse matrix. In our specific case, this distance constraint cannot be applied, thereby making a direct implementation of NCut essentially intractable.

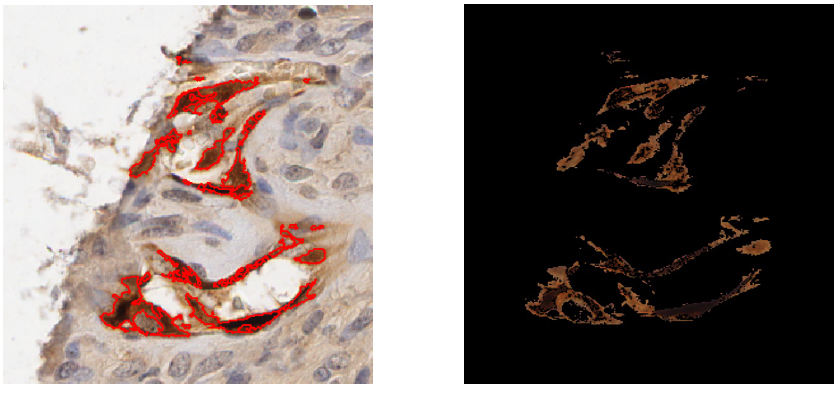


Fig. 1. The input (left) with annotated ground truth in red. Notice the difficulty in spotting stained cells with the untrained naked eye. Segmentation results (right) from HNCut.

2 Novel Contributions of This Work

The major contribution of this work is a novel, hierarchical unsupervised segmentation method termed HNCut. Additionally, to the best of our knowledge the application of HNCut to precisely quantify a vascular marker on OCa TMAs represents the first attempt at creating a quantitative, image based metric for OCa prognosis and survival. The novelty of HNCut is the use of a hierarchically represented data structure that bridges MS and NCut. The first component of HNCut, namely MS, efficiently prunes down the total number of clusters for the second component NCut. Unlike traditional clustering algorithms, our final aim is to extract a single cluster pertaining to the stained region (Fig. 1), while ignoring pixels in all other clusters. In traditional algorithms the pixels that normally should be removed are usually allocated to the least dissimilar cluster. In contrast, our setup encourages cuts that confidently trim away these undesired pixels. The unsupervised aspect of HNCut is particularly desirable for image analysis applications in histopathology where obtaining annotated samples for training a supervised classifier depends on the annotations provided by an expert and hence difficult to obtain. To summarize, the primary methodological contributions of this work are,

- A novel hierarchical segmentation approach (HNCut) that marries MS and NCut. The affinity matrix can now take advantage of multiple features, and multiple color spaces efficiently across large windows sizes.

- Parameter insensitive segmentation for large images and the ability of HNCut to discriminate between regions with similar color values. The parameter for the Gaussian kernel in the affinity matrix of NCut is automatically computed. The parameters for MS are automatically adjusted based on the variance of the output.

3 Hierarchical Normalized Cut

As mentioned earlier, our method has two broad components consisting of MS and NCut. The intuition for these are first presented, followed by the complete details in Section 3.4.

3.1 Mean Shift for Reducing Color Space Cardinality

MS is used to detect modes in data using density gradient estimation. With an image defined as $\mathcal{C} = (C, f)$ where C is a 2D Cartesian grid of N pixels, $c \in C = (x, y)$ representing the Cartesian coordinates of a pixel, and f a color intensity function associated with c, we have the fixed point iteration update $\forall c \in C$ in MS as

$$f_{k+1}(c) \leftarrow \frac{\sum_{i=1}^{N} G_{\sigma_{\text{MS}}}(f_k(c) - f_k(c_i)) f_k(c_i)}{\sum_{i=1}^{N} G_{\sigma_{\text{MS}}}(f_k(c) - f_k(c_i))} \tag{1}$$

where a Gaussian function, G, with a bandwidth parameter σ_{MS}, is used to compute the kernel density estimate at data point c. $k \in \{1, \ldots, K\}$ represents various levels of color resolution produced at each iteration. The overall computation time for Eq. 1 is $O(N^2)$. By employing the improved Fast Gauss Transform [6], we can reduce the complexity to $O(N)$ with minimal precision loss.

MS produces a feature-based pyramidal scene representation $\mathcal{C}_u = (C, f_u)$, where $u \in \{1, \ldots, K\}$ represent levels of the pyramid of height K. This results in a series of scenes $\mathcal{C}_u$, all mutually aligned, but with a fewer number of colors in the lower levels of the pyramid compared to the top. It is the fewer colors at the lower levels that enable NCut to be tractable, however, the upper levels of the pyramid are needed for high quality segmentation.

3.2 Normalized Cuts on Mean Shift Reduced Color Space

By setting the vertices (V) to the basins of attraction from the MS (i.e., unique color values), NCut can segment the data by representing it as a connected graph ($\mathbf{G}$=(E, V)), with edges (E) representing affinity or strength of connectivity. A cut is the processes by which the removal of edges leads to two partitions. A value of a normalized cut between two disjoint sets A and B is computed using

$$\text{NCut}(A, B) = \frac{\text{cut}(A, B)}{\text{assoc}(A, V)} + \frac{\text{cut}(A, B)}{\text{assoc}(B, V)} \tag{2}$$

where $\text{cut}(A,B) = \sum_{c_i \in A, c_j \in B} w(f(c_i), f(c_j))$, $i,j \in \{1,\ldots,N\}$. Additionally, $\text{assoc}(A,V) = \sum_{c_i \in A, c_k \in V} w(f(c_i), f(c_k))$, where $i,k \in \{1,\ldots,N\}$. Also, $A \cup B = V$ and $w(f(c_i), f(c_j))$ is a function used to compute an affinity measure between c_i and c_j. Normally, in NCut, an additional spatial constraint is introduced such that if $||c_i - c_j|| \geq \theta$,where θ is a spatial radius threshold, $w(f(c_i), f(c_j))$ is set to 0. Because of this spatial constraint, the affinity matrix $\mathbf{W}$ is normally sparse, making the method tractable. The goal of NCut is to identify partitions A, B such that the they have the most intragroup similarity, and the most intergroup dissimilarity. This process can be recast [8] and solved as a generalized eigenvalue system.

In our implementation of NCut we are concerned with partitioning the color space and not the image pixels per se. Hence if the process of MS in Section 3.1 on $\mathcal{C}$ eventually results in M unique colors, V is the set of vertices corresponding to these colors. Thus $\mathbf{W} \in \mathbb{R}^{\mathbf{M}\times\mathbf{M}}$ in the first application of NCut.

3.3 Integrating Domain Knowledge to Guide Normalized Cuts

Our goal with NCut is to discard A or B. We therefore define a swatch (color template) reflecting the attributes of the object of interest in the scene. We define $Q = \{f_{\alpha_1}, f_{\alpha_2}, \ldots, f_{\alpha_t}\}$ to represent the colors of the objects we seek. Note that Q is trivially determined by annotating (manually) a few pixels from the object of interest on a representative image and may be easily changed based on the application. Hence HNCut is still an unsupervised algorithm.

3.4 The HNCut Algorithm

Reducing color space cardinality with MS enables NCut to be tractable (irrespective of the distance constraint). However, the first application of NCut resulting in the sets A and B leaves us with too many false positives. We can overcome this draw back by maintaining a pyramid as we perform MS, as detailed below.

Step 1: Generate a hierarchical color pyramid via MS: For any scene $\mathcal{C}$, corresponding scenes in the pyramid $\{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$ of monotonically increasing resolutions $\{M_1, M_2, \ldots, M_K\}$ are generated.

Step 2: Apply NCut on lowest level of pyramid: Apply NCut on $\mathcal{C}_K$ to partition the scene into two color sets A_K and B_K. Use in the affinity matrix the values $\mathbf{W}_K(\beta_1, \beta_2) = \exp(-\frac{||\beta_1 - \beta_2||^2}{\sigma})$ where σ is a scaling parameter.

Step 3: Use color swatch to identify unique color partition: Identify which of A_K and B_K uniquely contains colors in the swatch Q. If $Q \subset A_K$ and $Q \cap B_K = \emptyset$ then eliminate B_K. If $Q \subset B_K$ and $Q \cap A_K = \emptyset$, similarly eliminate A_K. However if $Q \cap A_K \neq \emptyset$ and $Q \cap B_K \neq \emptyset$ then set $\sigma = \sigma + \Delta\sigma$ and repeat NCut on V_K to obtain a new partition. Keep incrementing σ until Q is uniquely contained within either of A_K or B_K.

Step 4: Repeat color partitioning at a fixed scale using NCut: Assuming for some value of σ, Q is uniquely contained in A_K, set $V = A_K$ and repeat

NCut on V. Repeat steps 2 and 3 until no further partitioning of color space at scale $\ell = K$ is possible.

Step 5: Repeat NCut and color partitioning at higher image scales: Migrate to the next higher image resolution, First, identify V as the set of colors at resolution $\ell = K - 1$ but derived from the non-decomposable partition $\ell = K$. Repeat steps 2-4 to identify a unique non-reducible color partition at $\ell = K - 1$ that contains the swatch. Also, do this step for subsequent lower values of $\ell \in \{1, \ldots, K - 2\}$

4 Experiment Setup and Methodology

Our database comprises of a total of seven digitized TMAs of OCa, in turn comprising over 500 tissue cylinders. The TMAs were obtained by sampling OCa tissue from over 100 patients and were stained for the presence of the TVM ESM-1, resulting in vascular regions with the antibody to ESM-1 staining brown. The TMA files are down-sampled and stored at 20x magnification, producing images that are approximately 1300 x 1400. An expert pathologist annotated 100 cylinders. The exact regions highlighted by him were extracted and stored separately, creating a binary map used as the ground truth.

Performance Metrics: Two different metrics were used to grade the performance of the algorithms. A macro metric (region) matched minimum surface area ellipses fit to both the test method and the ground truth to assess the correctness of region matching. A micro metric (pixel) was used to determine the amount of overall correctness of pixel matching.

4.1 Comparison of HNCut to *k*-Means, PBT

All algorithms loaded the same image and converted it to YCbCr space. To ensure fair comparisons, each algorithm operated only with chromatic information.

***k*-means.** Ten clusters were employed. The initial cluster centers were manually chosen, by experimentation, to provide a good representation of the image.

Probabilistic Boosting Tree. PBT was implemented as described in [3], using suggested default values for both θ and ϵ (.45 and .4, respectively). Each strong Adaboost classifier was constructed using seven weak classifiers. The PBT performed seven levels of dataset separation, in the prescribed tree like manner. The training set was created by taking a 3×3 window around every $c \in C$, across all 3 color channels in YCbCr space, resulting in a 27 dimensional vector. 1000 random positive (stained) samples and 1000 random negative (unstained and spuriously stained) samples were selected from 25 randomly selected images, resulting in a total training vector of size 27x50,000. 50 cross validation iterations took place resulting in an average AUC of .9296 with a standard deviation of .0039. The probability returned by the PBT was converted into a strong classifier by taking the upper Otsu threshold.

HNCut. MS was performed using a $\sigma_{MS} = .05$, with the number of clusters for the improved fast Gauss transform set to 350. NCut uses the Silverman function [10] to determine its initial σ and then increases by a factor of 10 as prescribed above. The domain knowledge is six pixels of shades of brown that we find acceptable as stain.

5 Results and Discussion

The first column in Fig. 2 represents the original input image, with the boundary of the ground truth highlighted by the pathologist labeled in red. The first row illustrates a case where all of the algorithms performed as expected. The second row illustrates an example where the HNCut algorithm performs optimally, while both the PBT and k-means extract many miscellaneous false positives. The hierarchical set of operations in HNCut is essentially a feedback loop that enables the algorithm to operate extremely efficiently and accurately. The final row is used to illustrate a scenario where false negatives (FN) occur for all three methods. The middle region is correctly segmented in all algorithms, while the three other regions are incorrectly rejected. The stain in those regions

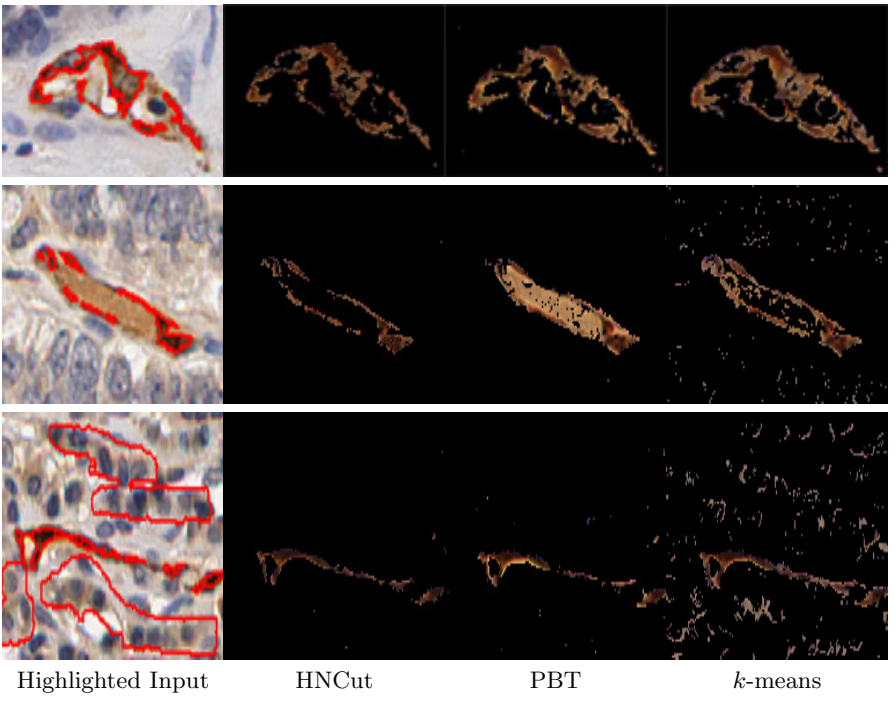


Fig. 2. Selected examples are presented. The original input, with the annotated ground truth in red, is presented on the left, followed by HNCut, PBT and k-means.

Table 1. Performance measures for HNCut, PBT and k-means

	Region					Pixel	
	FN	TP	FP	Sensitivity	PPV	Sensitivity	Specificity
HNCut	217	920	224	80.9%	80.4%	73.8%	99.7%
PBT	265	869	456	76.6%	65.6%	67.4%	99.6%
k-means	264	873	1646	76.8%	34.7%	69.1%	98.3%

is only barely visible to an expert. Lastly, we can see that although k-means does the best in attempting to detect those regions, the same setup results in many false positives in other images. This is a result of the k-means requiring all pixels to be assigned to a cluster, filling the stain cluster as it is the most similar out of the candidates. Both PBT and HNCut determine, in most instances, that these same values are simply too different to be considered as stain.

Table 1 quantitatively illustrates that HNCut provides the best balance between sensitivity and specificity, providing 50% less false positives (FP) than PBT and 86% less FP than k-means. All of the algorithms seem to do very well in the specificity category, but this is a result of the vast number of true negative pixels (TN) associated with the majority of the sample.

Using HNCut on 500 discs, about 10 of them failed to converge properly, resulting in very poor segmentations. Interestingly, these 10 images all had little to no stain present. This failure of convergence is due to the σ_{MS} for the MS being inappropriately selected, in all experienced cases being too wide. The automatic adjustment of the σ_{MS} took place, as described above, to a narrower setting resulting in appropriate results in all of the experiments we ran.

6 Concluding Remarks

In this paper we have presented a novel unsupervised segmentation scheme called Hierarchical Normalized Cuts (HNCut). The strength of HNCut is derived from the fact that it integrates the best of both the Mean Shift and the Normalized Cut algorithms. The presented algorithm is not only efficient and accurate but not encumbered by the need for precisely annotated training data. While HNCut could be applied to any general segmentation task, in this paper we evaluate it in the context of identifying a vascular biomarker in OCa TMAs. Results on 100 images reveal that HNCut significantly outperforms both Probabilistic Boosting Trees and k-means. The algorithm does use a color swatch but this involves manual identification of half a dozen pixels to represent features of interest. The swatch can be changed to suit similar biomedical segmentation problems swiftly. The swatch does not require laborious, careful annotation by an expert, but is a small sampling of the target class and may be provided by any user who is able to partially identify the target class.

References

1. Doyle, S., Madabhushi, A., Feldman, M.D., Tomaszeweski, J.E.: A boosting cascade for automated detection of prostate cancer from digitized histology. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 504–511. Springer, Heidelberg (2006)
2. Buckanovich, R., Sasaroli, D., et al.: Tumor vascular proteins as biomarkers in ovarian cancer. Journal Of Clinical Oncology, 852–861 (March 2007)
3. Tu, Z.: Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering. In: ICCV, pp. 1589–1596 (2005)
4. Fukunaga, K., Hostetler, L.: The estimation of the gradient of a density function, with applications in pattern recognition. IEEE Trans. Information Theory 21(1), 32–40 (1975)
5. Comaniciu, D., Meer, P.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. PAMI 24(5), 603–619 (2002)
6. Yang, C., Duraiswami, R., Gumerov, N., Davis, L.: Improved Fast Gauss Transform and efficient kernel density estimation. In: ICCV, pp. 664–671 (2003)
7. Wu, Z., Leahy, R.: An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. IEEE Trans. PAMI 15(11), 1101–1113 (1993)
8. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. PAMI 22(8), 888–905 (2000)
9. Dhillon, I.S., Guan, Y., Kulis, B.: Weighted graph cuts without eigenvectors a multilevel approach. IEEE Trans. PAMI 29(11), 1944–1957 (2007)
10. Silverman, B.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, Boca Raton (1986)

Nonparametric Intensity Priors for Level Set Segmentation of Low Contrast Structures

Sokratis Makrogiannis[1], Rahul Bhotika[1], James V. Miller[1], John Skinner Jr.[2], and Melissa Vass[2]

[1] GE Global Research, One Research Circle, Niskayuna, NY 12309, USA
{makrogia,bhotika,millerjv}@research.ge.com
[2] GE Healthcare, 3000 N. Grandview Blvd., Waukesha, WI 53188, USA
{melissa.vass,john.skinner}@ge.com

Abstract. Segmentation of low contrast objects is an important task in clinical applications like lesion analysis and vascular wall remodeling analysis. Several solutions to low contrast segmentation that exploit high-level information have been previously proposed, such as shape priors and generative models. In this work, we incorporate *a priori* distributions of intensity and low-level image information into a nonparametric dissimilarity measure that defines a local indicator function for the likelihood of belonging to a foreground object. We then integrate the indicator function into a level set formulation for segmenting low contrast structures. We apply the technique to the clinical problem of positive remodeling of the vessel wall in cardiac CT angiography images. We present results on a dataset of twenty five patient scans, showing improvement over conventional gradient-based level sets.

1 Introduction

This paper introduces a method for segmenting low contrast regions in Computed Tomography (CT) volumes by integrating local nonparametric intensity statistics into the level set framework. The drive for early detection and quantification of disease has greatly improved the spatial resolution and sensitivity of CT scanners. This has led to the growing need for methods for sub-pixel accurate segmentation and boundary delineation of small structures with volumes ranging from 25 mm^3 to 500 mm^3, or equivalently, 200 to 4000 voxels. Segmentation of such small structures is further complicated by low contrast, partial volume averaging, and other imaging artifacts that make it difficult to robustly detect the boundary between the object of interest and its background and surrounding structures (see Fig. 1). Furthermore, the underlying anatomical structures of interest often have intensity distributions that are not robustly captured by standard parametrizations or lower-order sufficient statistics, as shown in Fig. 2. These factors motivate the importance of continued investigation into methods for improving the segmentation of small, low-contrast structures with spatially varying intensity information.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 239–246, 2009.

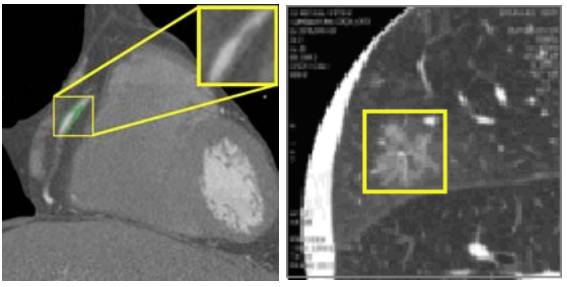


Fig. 1. Examples of low contrast structures in CT images. The image on the left shows a coronary artery with a thickened wall barely visible against the background cardiac tissue. The image on the right shows a subsolid nodule that appears "translucent" and diffuses into surrounding lung tissue.

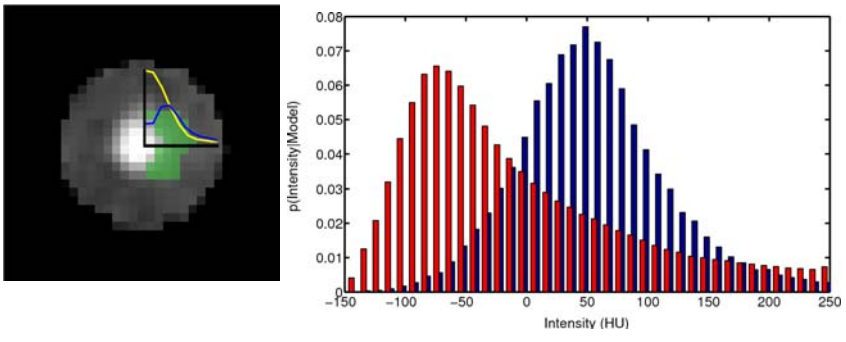


Fig. 2. Left: Example of ramp intensity profile (yellow) and corresponding gradient magnitude (blue) that causes gradient-based level set to propagate beyond the real boundary given by the reference segmentation (green). Right: Histograms of pixel intensities in vessel wall remodeling regions (blue) and in nearby cardiac tissue and background (red) were computed over 20 subjects. The histograms overlap considerably, but have overall distinctive distributions. Note also that the cardiac tissue distribution is not a Gaussian and difficult to parametrize.

Previously, approaches using parametric deformable surface models that propagate by minimizing an energy functional which includes internal (stretching, bending), image (usually gradient-based), and external (user-defined or derived from priors) forces have been proposed [1,2]. These often produce accurate segmentation but have limitations with respect to topological flexibility and the segmentation of complex structures. These limitations are somewhat overcome by level set methods, which represent the evolving surface implicitly. Early techniques such as fast marching supported only monotonic motion [3] and were followed by more sophisticated approaches such as shape-based segmentation [4], geodesic active contours [5], or methods using prior shape statistics to regularize the level set formulation [6,7]. A solution for the piecewise constant case of the Mumford-Shah functional [8] in the level set context was presented in [9],

but only accounted for first-order foreground and background statistics, approximated using average intensities. In [10], a data consistency term was used to incorporate local intensity information, but the method makes several assumptions: the overall image intensity distribution is assumed to be bimodal such that it can be factored into underlying foreground and background distributions which themselves are assumed to be unimodal, and whose parameters can be recovered from an initial coarse segmentation. Recently, the Bhattacharyya distance, a nonparametric dissimilarity measure, was integrated into the level set framework [11], but the method does not allow the use of prior statistics to drive the segmentation process.

This paper introduces a technique that makes use of nonparametric prior models of intensity within the level set variational framework to accurately segment low contrast regions. Reference segmentations provided by clinical specialists were used to build nonparametric models of the intensity distribution of the object of interest and its surrounding background. During segmentation, we first estimate the dissimilarity of the local intensity distribution in the vicinity of each voxel in the image to the two model distributions using Earth Mover's Distance (EMD) or the Mallows Distance [12,13] as a local indicator function for the likelihood of the voxel belonging to the foreground. The local indicator function is integrated with gradient-based information to define a new speed term in the level set partial differential equation. There are two major contributions of this work. First, it incorporates and balances both boundary and texture information into the segmentation process, while retaining the topological flexibility, sub-pixel accuracy, and boundary extraction capabilities of the level set framework. Second, it uses prior statistical knowledge about the object and background intensity distributions without imposing a specific parametrization, e.g., a Gaussian distribution, on the class models. In the remainder of this paper, we provide the theoretical background, formulate our approach, present a clinical application, and show experimental results.

2 Background: The Level Set Formulation

In level set approaches the propagating surface $S(t)$ is represented by the zero level set of a higher dimensional function $\psi(\omega(t);t) : \mathbb{R}^3 \rightarrow \mathbb{R}$ that evolves with time in the Eulerian coordinate system, and may be defined as $S(t) \equiv \psi(\omega(t);t) = 0$. The main benefits of the level set formulation are topological flexibility and differentiability [3,4]. The basic equation of level set propagation is given by:

$$\frac{\partial \psi}{\partial t} + F(\omega(t);t) \cdot |\nabla \psi| = 0, \tag{1}$$

where $F(\cdot)$ is the speed function that controls the evolution of the level set and $\omega(t)$ denotes the spatial coordinates in the d-dimensional space $\omega \in N^d$ that varies with time t. Given an initial position of the surface, $S(t = 0)$, a signed distance transform is typically applied to generate $\psi(\omega(t = 0);t = 0)$. After setting the initial conditions, the evolution of $\psi(\omega(t);t)$ and deformation of $S(t)$ are obtained by solving (1) at each time point t.

In the conventional level set framework, surface evolution is governed by geometric and image-based features, such as the surface curvature κ, the gradient magnitude of the image intensity I, and an external (inflation/deflation) driving force c [4]. This can be expressed as

$$F_G(\omega(t);t) = [c + \kappa(\omega(t);t))] \cdot g(\nabla I) \cdot \varsigma(P), \tag{2}$$

where $g(\nabla I) = 1/\left(1 + |\nabla I * G_\sigma|\right)$ with G_σ being a Gaussian with standard deviation σ used to smooth out noise. P is an optional spatial prior term that may be used if prior shape, size, or location information is available. Also, $\varsigma(X)$ is the sigmoid function that normalizes the dynamic range of a feature X and is given as $\varsigma(X) = X_{\min} + (X_{\max} - X_{\min})/\left(1 + \exp\left(-\frac{X-\beta}{\alpha}\right)\right)$. When the boundary between the object and background is diffuse, as is often the case in biomedical imaging, the edge profile follows a ramp model (see Fig. 2 (left)). In these cases, gradient-based implementations as in (2) tend to propagate over the boundary, leaking into the background. More accurate segmentation can be achieved by using prior statistical information about the intensity distribution of an object and its surrounding area to complement the gradient information. However, first and second order statistics or parametric representations of distributions are often not sufficient for separation of the object from the background, as shown in Fig. 2 (right). We propose the integration of higher order intensity statistics into the level set formulation via a nonparametric dissimilarity measure for segmenting low contrast structures.

3 Incorporating Nonparametric Priors in Level Sets

A local indicator function, defined at each voxel and denoted by $H(\omega)$, that expresses the likelihood of the voxel belonging to a foreground object can be integrated in the level set formulation along with gradient magnitude and curvature to define a new speed term F_H as follows (with the help of (2)):

$$F_H(\omega(t);t) = F_G(\omega(t);t) \cdot \varsigma(H), \tag{3}$$

where $\varsigma(\cdot)$, again, is the sigmoid function. So far, the only requirement imposed on $H(\cdot)$ is that it can be evaluated locally. Next, we define and derive a form of $H(\cdot)$ in terms of a dissimilarity measure of a locally sampled intensity distribution to nonparametric intensity distributions of foreground and background classes established from exemplars.

We represent the prior statistical information of foreground and background regions using normalized histograms of intensity. The histogram of foreground regions was learned from clinical image datasets where the foreground regions were delineated by clinical imaging experts. The histogram of the background was derived from the same datasets by automatically selecting regions surrounding the delineated foreground voxels. Fig. 2 (right) shows the histograms for remodeled sections of coronary vessel wall and the nearby cardiac tissue over a set of twenty CT volumes.

The histograms of the foreground and background intensity models are used to construct the indicator function $H(\cdot)$ in (3). At each voxel ω, a histogram built from a small neighborhood about ω is compared to the model foreground and background histograms using the Earth Mover's Distance [12,13] as a well defined norm on the differences between two distributions

$$D^2(\omega; M_i, \mathcal{N}(\omega)) = \sum_k \left[\mathrm{CDF}(k, M_i) - \mathrm{CDF}(k, \mathcal{N}(\omega))\right]^2, \tag{4}$$

where M_i is a model (foreground or background), $\mathcal{N}(\omega)$ is a neighborhood about ω, and $\mathrm{CDF}(\cdot)$ is the cumulative probability distribution of intensity for a model or over a neighborhood. The dissimilarity measures of the sample histogram over $\mathcal{N}(\omega)$ to each model histogram are combined into an indicator function given by

$$H(\omega) = D^2(\omega; M_1, \mathcal{N}(\omega)) - D^2(\omega; M_2, \mathcal{N}(\omega)) - T \tag{5}$$

where M_1 and M_2 denote the background and object model respectively, and T represents a difference between the prior probabilities of the two models. $H(\cdot)$ can be plugged into (3) to give the nonparametric intensity-distribution weighted speed function, F_H.

4 Clinical Application: Vessel Wall Segmentation

The proposed algorithm was tested on cardiac CT angiography (CTA) datasets in the context of vessel wall delineation for measuring positive remodeling. CTA is a minimally invasive imaging procedure used to examine the health of blood vessels in key areas of the human body such as brain, heart, lung, abdomen, and kidneys. It involves injection of a contrast agent, usually iodine-based, to enhance the image contrast inside the vessel lumen. This enables the segmentation of the lumen using simple algorithms such as morphological operations or region growing and is useful for detecting stenoses, i.e., narrowing of the vessel lumen.

Positive wall remodeling, on the other hand, is defined as an increase in the vessel wall thickness, resulting in the bulging out of the outer boundary of the vessel wall. This is usually observed at atherosclerotic lesion sites and often indicates an unstable clinical presentation associated with high risk of an acute cardiac event [14,15]. Accurate segmentation plays a significant role here because the coronary vessel walls are thin, typically 2-5 pixels wide in state-of-the-art CT. Unfortunately, compared to the vessel lumen, the image contrast of the vessel wall is quite low. This is due to surrounding cardiac tissue, partial voluming, and noise. Thus, accurate localization of the wall boundary using gradient-based level sets is not possible. On the other hand, the intensity distribution of the positive remodeling voxels marked by clinical specialists was observed to differ from that of its surrounding tissue and background, as shown earlier in Fig. 2. Therefore, the proposed speed function given in (3) which incorporates nonparametric intensity statistics should yield improved segmentation.

5 Experimental Results

Preprocessing steps that are not the focus of this paper were used to isolate the bright contrast-filled vessel lumen regions in CTA datasets. For evaluating the proposed algorithm, we operated only on the vessel segments containing examples of wall remodeling. A volume equal to three times the nominal diameter of the vessel lumen was fixed as the domain for the level set propagation. A signed distance function was computed from the lumen surface and the level set was initialized at 0.5 mm outside the lumen.

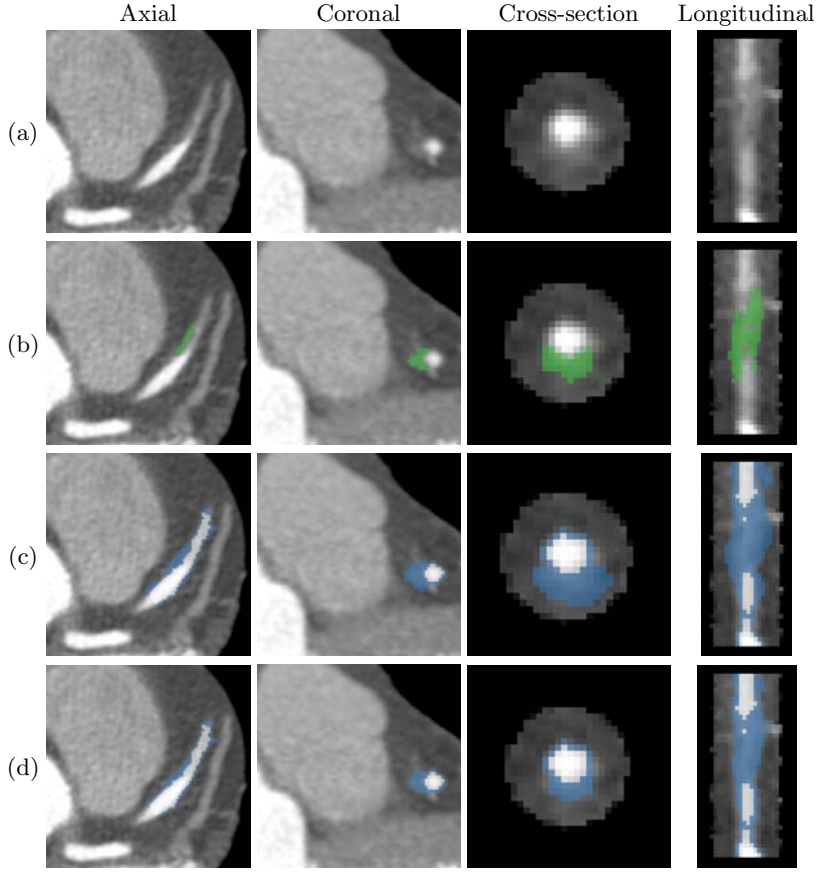


Fig. 3. Segmentation results. Row (a) Original CT intensity data, (b) reference standard showing positive remodeling regions overlaid on the CT data, (c) segmentation using conventional gradient-based level sets (G-LS), and (d) segmentation results using the proposed histogram-based level set (H-LS) approach. The first two columns show axial and coronal views. For better visualization, the last two columns show reformatted views across and along the vessel. G-LS propagates over the wall boundary into the background areas, while H-LS produces more accurate delineation.

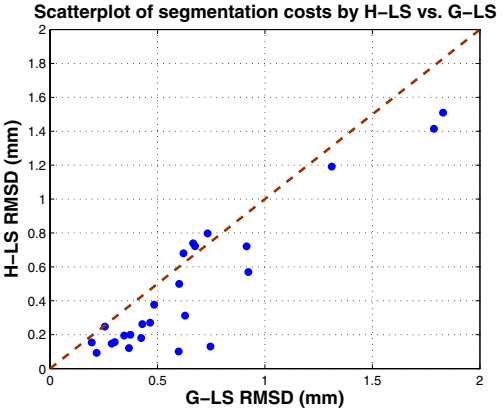


Fig. 4. Scatterplot of segmentation costs produced by H-LS versus G-LS for 25 patient CT scans (in blue). The segmentation cost is the root mean squared distance (RMSD) of the segmented surface from the reference surface. Points below the diagonal red line correspond to lower RMSD values of H-LS compared to G-LS, i.e., better segmentation accuracy. We observe that H-LS outperforms G-LS in 21 out of 25 cases.

For the sake of brevity in the ensuing discussion, the level set scheme using the proposed speed function (3) is referred as "H-LS" and the one using the conventional speed function (2) as "G-LS". Results for both algorithms were evaluated against reference standard segmentations of positive remodeling wall regions that were provided by clinical specialists. Representative segmentation results are illustrated in Fig. 3 showing H-LS gives segmentations much closer to the reference positive remodeling regions, while G-LS tends to leak and propagate beyond the vessel wall boundary. We use the root mean squared distance (RMSD) of the segmented surface from the surface of the reference segmentation to quantify performance. We add the vessel lumen to both the reference and segmented volumes prior to calculating RMSD to mitigate large distances from false vessel wall detections opposite to the actual positive wall remodeling. Fig. 4 compares the RMSD values for G-LS and H-LS over twenty five datasets. The proposed H-LS method consistently outperforms standard G-LS. The mean RMSD over the twenty five cases was found to be 0.65 mm for G-LS and 0.47 mm for H-LS, giving a 28% improvement on average. Furthermore, a paired t-test between the RMSDs produced by H-LS and G-LS gives a p-value of 0.000032 suggesting statistically significant improvement.

6 Conclusion

We introduced a method for naturally incorporating nonparametric intensity priors of foreground and background objects into a local, nonparametric dissimilarity measure as a speed function term in the level set framework. The method improves segmentation accuracy for low contrast regions by providing a counterbalance to gradient-based front propagation. We demonstrated the method on a

database of vessels with positive wall remodeling. While vessel lumen narrowing or stenosis can be detected by directly analyzing the vessel lumen, our technique allows for the accurate segmentation of the remodeled wall region, a necessary step for quantifying pathology. Our approach is applicable to other low contrast segmentation problems where exemplars are available. The low-level segmentation scheme can also be viewed as an independent step or combined with additional analysis and classification approaches to aid in diagnosis.

References

1. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1(4), 321–331 (1988)
2. Xu, C., Prince, J.L.: Snakes, shapes, and gradient vector flow. IEEE Transactions on Image Processing 7(3), 359–369 (1998)
3. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. In: Proceedings of the National Academy of Sciences, pp. 1591–1595 (1996)
4. Malladi, R., Sethian, J.A., Vemuri, B.C.: Shape modeling with front propagation: A level set approach. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(2), 158–175 (1995)
5. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. International Journal of Computer Vision 22(1), 61–79 (1997)
6. Leventon, M., Faugeras, O., Grimson, W.: Level set based segmentation with intensity and curvature priors. In: Workshop on Mathematical Methods in Biomedical Image Analysis, pp. 4–11 (2000)
7. Paragios, N., Deriche, R.: Geodesic active regions and level set methods for supervised texture segmentation. International Journal of Computer Vision 46(3), 223–247 (2002)
8. Mumford, D., Shah, J.: Optimal approximations by piecewise smooth functions and associated variational problems. Communications on Pure and Applied Mathematics 42(5), 577–685 (1989)
9. Chan, T., Vese, L.: Active contours without edges. IEEE Transactions on Image Processing 10(2), 266–277 (1999)
10. Baillard, C., Barillot, C.: Robust 3D segmentation of anatomical structures with level sets. In: Medical Image Computing and Computer-assisted Intervention, pp. 236–245 (2000)
11. Rathi, Y., Michailovich, O., Malcolm, J., Tannenbaum, A.: Seeing the unseen: Segmenting with distributions. In: International Conference on Signal and Image Processing (2006)
12. Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. International Journal of Computer Vision 40(2), 99–121 (2000)
13. Levina, E., Bickel, P.: The earth mover's distance is the Mallows distance: some insights from statistics. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 251–256 (2001)
14. Glagov, S., Weisenberg, E., Zarins, C.K., Stankunavicius, R., Kolettis, G.J.: Compensatory enlargement of human atherosclerotic coronary arteries. New England Journal of Medicine 316(22), 1371–1375 (1987)
15. Crawford, M., DiMarco, J.P., Paulus, W.J. (eds.): Cardiology. Elsevier, Amsterdam (2004)

Improving Pit–Pattern Classification of Endoscopy Images by a Combination of Experts

Michael Häfner[1], Alfred Gangl[1], Roland Kwitt[2], Andreas Uhl[2], Andreas Vécsei[4], and Friedrich Wrba[3]

[1] Dept. of Gastroenterology & Hepatology, Medical University of Vienna, Austria
[2] Dept. of Computer Science, University of Salzburg, Austria
{rkwitt,uhl}@cosy.sbg.ac.at
[3] Dept. of Clinical Pathology, Medical University of Vienna, Austria
[4] St. Anna Children's Hospital, Vienna, Austria

Abstract. The diagnosis of colorectal cancer is usually supported by a staging system, such as the Duke or TNM system. In this work we discuss computer–aided pit–pattern classification of surface structures observed during high–magnification colonoscopy in order to support dignity assessment of colonic polyps. This is considered a quite promising approach because it allows in vivo staging of colorectal lesions. Since recent research work has shown that the characteristic surface structures of the colon mucosa exhibit texture characteristics, we employ a set of texture image features in the wavelet-domain and propose a novel classifier combination approach which is similar to a combination of experts. The experimental results of our work show superior classification performance compared to previous approaches on both a two-class (non-neoplastic vs. neoplastic) and a more complicated six-class (pit–pattern) classification problem.

1 Motivation

Recent statistics of the American Cancer Society reveal that colorectal cancer is the third most common cancer in men and women and the second most common cause of US cancer deaths. Since most colorectal cancers develop from polyps, a regular inspection of the colon is recommended in order to detect lesions with a malignant potential or early cancer. A common medical procedure to examine the inside of the colon is colonoscopy, which is usually carried out with a conventional video–endoscope. A diagnostic benefit can be achieved by employing so called high–magnification endoscopes (aka zoom–endoscopes), which achieve a magnification factor of up to 150 by means of an individually adjustable lens. In combination with dye–spraying to enhance the visual appearance (chromo–endoscopy) of the colon mucosa, high–magnification endoscopy can reveal characteristic surface patterns, which can be interpreted by experienced physicians. Commonly used dyes are either methylene-blue, or indigo–carmine, which both lead to a plastic effect. In the research work of Kudo et al. [1], the macroscopic appearance of colorectal polyps is systematically described and results in the so called *pit–pattern* classification scheme.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 247–254, 2009.

The contribution of this work is a novel way for classifier combination to enhance the accuracy of differential diagnosis. We propose a fusion of three approaches from classification research and show that by using several recently proposed texture image features for endoscopy image classification, we achieve a remarkable increase in overall classification accuracy.

The remainder of this paper is structured as follows: in Sect. 2, we review the medical background an introduce the pit–pattern classification scheme. Section 3 discusses the feature extraction step as well as our classification approach. In Sect. 4, we present the experimental results of our work and Sect. 5 concludes the paper with a short summary and an outlook on future research.

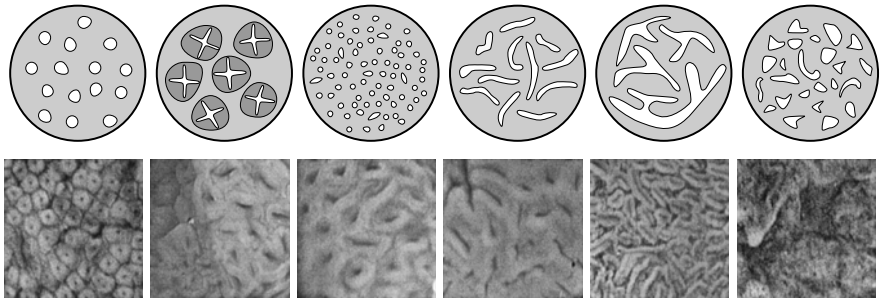
2 Pit-Pattern Classification

Polyps of the colon are a frequent finding and are usually divided into metaplastic, adenomatous and malignant. Since the resection of all polyps is rather time–consuming, it is imperative that those polyps which warrant resection can be distinguished. Furthermore, polypectomy[1] of metaplastic lesions is unnecessary and removal of invasive cancer may be hazardous. The classification scheme presented in [1] divides the mucosal crypt patterns into five types (pit–patterns I–V, see Fig. 1), which can be observed using a high–magnification endoscope.

While types I and II are characteristic of benign lesions and represent normal colon mucosa or hyperplastic polyps (non–neoplastic lesions), types III to V represent neoplastic, adenomatous and carcinomatous structures. Our classification problem can be stated as follows: the problem of differentiating pit–patterns I and II from III–L,III–S,IV and V will be denoted as the *two–class* problem (non–neoplastic vs. neoplastic), whereas the more complex and detailed discrimination of all pit–patterns I to V will be denoted as the *six–class* problem. At first sight, the pit–pattern classification scheme seems to be straightforward and easy to be applied. Nevertheless, it needs some experience and exercising to achieve good results. Correct diagnosis very much relies on the experience of the endoscopist as the interpretation of the pit–patterns may be challenging [2].

Our approach is motivated by the work of Kato et al. [3], where the authors state that assessing the type of mucosal crypt patterns can actually predict the histological findings to a very high accuracy. Regarding the correlation between the mucosal pit–patterns and the histological findings, several studies reported good results, although with quite different diagnostic accuracies. A comparative study by Kato et al. [4] shows that the classification accuracy in magnifying colonoscopy ranges from 80.6% to 99.1%. Another extensive study by Hurlstone et al. [5] report error rates of approximately 5%. In [6] the authors claim 95.6% for chromoendoscopy with magnification in contrast to diagnosis using conventional colonoscopy (84.0%) and diagnosis using chromoendoscopy without magnification (89.3%). In addition to that, inter–observer variability of magnification chromoendoscopy has been described at least for Barret's esophagus [7]. This inter–observer variability may to lesser degree be also present in

[1] The process of removing polyps.



(a) I (b) II (c) III-S (d) III-L (e) IV (f) V

Fig. 1. Schematic illustration of the pit–pattern characteristics (top row) together with exemplary pit–pattern images obtained during high–magnification colonoscopy (bottom row)

the interpretation of pit–patterns of colonic lesions. This work aims at allowing computer–assisted pit–pattern classification in order to enhance the quality of differential diagnosis.

3 Feature Extraction and Classification

We use a selection of texture feature extraction approaches in the wavelet–domain which have already been successfully applied in the context of endoscopy image classification. Our first feature set is computed using an approach presented in [8] where the authors decompose each image using the Dual–Tree Complex Wavelet Transform (DT–CWT) and model the absolute values of detail subband coefficients by two–parameter Weibull distributions. The Maximum–Likelihood estimates of the Weibull scale & shape parameter of each subband are then arranged into feature vectors for nearest–neighbor classification using the Euclidean distance. Although the work in [8] uses grayscale images, it can easily be extended for color images by simple feature vector concatenation of separately computed color–channel feature vectors. In another work [9], a set of image features is computed from both the classic pyramidal (DWT) and undecimated wavelet transform (SWT) by calculating so called color eigen–subband features (CES). The CES features are essentially the eigenvalues obtained during PCA on the stacked wavelet detail subbands of the color–channels. Since the study in [8] has shown that the DT–CWT produces highly discriminative features, we have extended the CES approach to work with the DT–CWT. The last feature extraction approach we take into account is presented in [10] and it based on the computation of so called color wavelet–energy correlation signatures (WECS) between wavelet–decomposed color–channels. The extension to the DT–CWT is again straightforward and the dimensionality of the feature vectors is doubled. In the following, feature vectors will be denoted by $\mathbf{v} \in \mathcal{F} \subset \mathbb{R}^d$, where $\mathcal{F}$ denotes the feature space and d denotes the feature space dimensionality.

3.1 Classification Setup

We propose a fusion of three main approaches from recent work in classification research. We combine nearest–neighbor (NN) classifiers in a One–Against–One [11] (aka round–robing or pairwise coupling) setup and optimize each classifier using sequential forward feature subset selection (SFFS) [12]. The different class predictions are then combined using a *voting–against* approach and class posterior probability estimation.

The One–Against–One Approach (OAO). The basic concept of the OAO classification strategy is to split a multi–class problem into smaller binary problems with the ulterior motive that the decision boundaries for binary problems are simpler and easier to learn than the decision boundaries in case of a multi–class problem. In OAO classification, one classifier is trained for each possible pair of classes. Given a c–class problem, we thus have to train a total of $c(c-1)/2$ classifiers. In the training stage, each binary classifier is trained using only the examples of the two classes it has to discriminate. For example, given that $V := \{\mathbf{v}_i\}_{1 \leq i \leq L}$ denotes the complete set of training vectors and $\gamma : \mathcal{F} \rightarrow \{1, \ldots, c\}$ denotes a function returning the true class membership of $\mathbf{v}$, then the binary classifier $C_{ij} : \mathcal{F} \rightarrow \{i, j\}$ – which discriminates between class i and j – is trained using the training subset $S_{ij} := \{\mathbf{v}_n | \gamma(\mathbf{v}_n) = i \vee \gamma(\mathbf{v}_n) = j\}$. Hence, we can view the OAO approach as some sort of expert system, where each classifier is an expert in discriminating only two particular classes. In such a classifier combining approach we require two important properties of the base classifiers [13]: *diversity* and *accuracy*. First, diversity signifies that the errors should be uncorrelated and second, accuracy refers to a classification accuracy of at least 50%. In pratical use, we also require efficiency, which refers to low computational cost. However, due to space limitations we will not deal with this issue here.

Increasing Diversity. The problem of combining NN classifiers is particularly interesting in the context of the diversity requirement, since the main approaches of classifier combination to increase diversity, such as bagging [14] or boosting [15] do not lead to the desired results. The root of the problem is the insensitivity of the NN classifier to changes in the training patterns, which is essentially the starting point for both bagging and boosting. In [16] this issue is discussed in great detail and a new random feature subset selection approach is proposed, where each classifier works on a random subset of all available features. Due to the sensitivity of the NN classifier w.r.t. changes in the feature set, this approach can increase diversity. Instead of using random feature subsets, we select each subset by means of SFFS, imposing no limit on the size of the resulting subsets. Starting with a subset of cardinality one, one feature is added in each iteration in case this feature improves leave–one–out crossvalidation accuracy. By using SFFS, we cover both requirements of accuracy and diversity at the same time. In combination with the OAO approach, we obtain $c(c-1)/2$ feature subsets after the training stage.

Combining Class Predictions. Since each classifier in our OAO ensemble will provide a class prediction, the question arises of how to combine the $c(c-1)/2$ predictions. Although it seems straightforward to employ a simple majority voting rule, this rule is logically incorrect w.r.t. OAO classification for one simple reason: given an arbitrary sample $\mathbf{v} \in \mathcal{F}$, a classifier C_{ij} will output either i or j as the predicted class label. However, this prediction is convenient, if and only if the sample $\mathbf{v}$ actually belongs to either class i or j. In that case the prediction is termed a *qualified* prediction. Otherwise, the prediction is termed an *unqualified* prediction. As a consequence, given that $C_{ij}(\mathbf{v}) = i$, we can at best conclude that $\mathbf{v}$ is not a member of class j. This interpretation is known as *voting against* [17] in contrast to *voting for*, which is correct only in case each classifier was trained to discriminate samples from all classes. The final prediction is obtained by counting the votes against each class and selecting the very one which received the smallest number of votes–against. Although the idea of voting–against seems to be pedantic at first sight, it allows a quite efficient way to compute the final prediction [11] and enables us to compute a closed–form estimation of the class posterior probabilities $P(i|\mathbf{v})$ [17]. Given that ϵ_{ji} is defined as the probability $P(C_{ij}(\mathbf{v}) = i|\gamma(\mathbf{v}) = j)$ (i.e. classifier C_{ij} outputs i though the sample belongs to class j) and w_i denotes the a–priori class probability of class i, the logarithm of the class posterior probability of class i can be calculated by

$$\begin{aligned} \log P(i|\mathbf{v}) = K + \log w_i + \\ \sum_{i \neq j} \log \left(\epsilon_{ji}, \text{ if } C_{ij}(\mathbf{v}) = j; \frac{1 - \epsilon_{ij} w_j}{1 - w_j} \text{ if } C_{ij}(\mathbf{v}) = i \right) + \\ \sum_{k, i \neq j} \log \left(\frac{1 - \epsilon_{kj} w_j}{1 - w_j} \text{ if } C_{kj}(\mathbf{v}) = k; \frac{1 - \epsilon_{jk} w_k}{1 - w_k} \text{ if } C_{kj}(\mathbf{v}) = j \right) \quad . \end{aligned} \tag{1}$$

The error terms ϵ_{ji} can be easily estimated in the training stage of the system from the outputs of classifier C_{ij} when presenting samples $\mathbf{v}$, $\gamma(\mathbf{v}) \neq i, j$. Further, the term K is simply a constant which is of no particular relevance for determining the final prediction. By using (1) we determine the predicted class label k (or equivalently the predicted in vivo staging of the endoscopy image) of a feature vector $\mathbf{v}$ by $k = \arg\max_i \log(P(i|\mathbf{v}))$.

4 Experimental Results

Our image database contains 484 RGB images of size 256×256, acquired in 2005/2006 at the Department of Gastroenterology and Hepatology (Medical

Table 1. Number of image samples per pit–pattern

I	II	III–L	III–S	IV	V
126	72	62	18	146	60

Table 2. Classification accuracy results for six different feature sets and the two combining approaches together with the McNemar–test results

	Voting–Against	**Cutzu [17]**		**Original Work**	
DWT & WECS [10]	90.08	91.53	+	79.96	++
DT–CWT & WECS [10]	95.04	95.25	–	86.57	++
DT–CWT & Classic [8]	96.69	97.11	–	93.18	++
DT–CWT & Weibull [8]	97.11	97.31	–	94.01	++
DT–CWT & CES [9]	97.31	97.73	–	88.84	++
DWT & CES [9]	93.18	94.43	+	84.09	++

University of Vienna) using a magnification endoscope (Olympus Evis Exera CF–Q160ZI/L) with a magnification factor of 150x. To enhance visual appearance, dye–spraying with indigo–carmine was applied and biopsies or mucosal resections were taken to obtain a histopathological diagnosis (*our ground truth*). For pit–patterns I,II and V, biopsies were taken, since these types need not be removed. Lesions of pit–pattern types III–S,III–L and IV have been removed endoscopically. Table 1 lists the number of image samples per class. We use exactly the same setup for feature extraction as presented in the original works [8,9,10], discussed in Sect. 3. The maximum decomposition depth of the wavelet transforms is set to $J = 6$. Regarding the dimensionality d of the resulting feature spaces $\mathcal{F}$ (using the DT–CWT), we obtain $d = 18J$ for [9], $d = 36J$ for [8] and $d = 18J$ for [10]. In case the DWT is used for the WECS approach, we obtain $d = 9J$. Table 2 lists the maximum leave–one–out crossvalidation accuracies for all feature extraction approaches and the two classifier combining schemes compared to the highest accuracies achieved in the original (color–extended) works. Since most of the results – especially between the combining schemes – are very similar, we conduct a McNemar–test [18] to test for statistically significant differences at the 5% significance level. The null–hypothesis H_0 is that there is no significant difference. A '+' indicates a rejection of H_0, while a '−' indicates that H_0 could not be rejected. Column four of Table 2 lists the McNemar–test results when comparing the combining schemes, column six lists the results when comparing the original work to the voting–against (first ± entry) and class posterior probability estimation (second ± entry) approach. As we can see, the best overall accuracy is obtained by the DT–CWT & CES features with 97.73%. We further notice that in the majority of cases, there is no significant difference between direct voting–against and class posterior probability estimation. However, compared to the original works, the OAO results are significantly superior with an average increase in leave–one–out crossvalidation accuracy of $\approx 8\%$. To get an impression of the misclassifications per class, Table 3 shows the confusion matrix of the DT–CWT & CES result. By breaking down the six–class problem to the two–class problem (see Sect. 2) we obtain an overall leave–one–out accuracy of 99.59%, which is considerably higher than in the original works. As a last note, we remind that although we use leave–one–out crossvalidation, all reported accuracies are actually *training* accuracies. Since high–magnification endoscopy is a rather new method for the diagnosis of colorectal cancer, there exists a

Table 3. Detailed confusion matrix results for the DT–CWT & CES features using Cutzu's class posterior probability estimation

	I	**II**	**III–S**	**III–L**	**IV**	**V**
I	**123**	3	0	0	0	0
II	4	**67**	0	0	1	0
III–S	0	0	**62**	0	0	0
III–L	0	0	0	**18**	0	0
IV	1	0	0	0	**145**	0
V	0	0	0	0	2	**58**

lack of data material which prevents to separate an independent set of test–images. As a result, it is highly probable that the accuracies are overestimated in a sense. Nevertheless, our results clearly indicate that computer–assisted pit–pattern classification based on the visual appearance of the colon mucosa can predict the histological results to a large extent.

5 Conclusion

In this paper[2], we have exploited the idea of combining a number of two–class classifiers to obtain an diagnostic prediction for high–magnification colonoscopy images. Our results show a remarkable improvement in leave–one–out crossvalidation accuracy compared to previous works. Since most of the computational effort (mainly feature selection) resides in the training stage there is no limiting factor which might prevent practical application. Depending upon the availability of a larger dataset, future research includes an evaluation of the approach using clearly separated training and test sets which is currently impossible.

References

1. Kudo, S.: Colorectal Tumours and Pit Pattern. Journal of Clinical Pathology 47(10), 880–885 (1994)
2. Hurlstone, D.: High-Resolution Magnification Chromoendoscopy: Common Problems Encountered in Pit-Pattern Interpretation and Correct Classification of Flat Colorectal Lesions. American Journal of Gastroenterology 97(4), 1069–1070 (2002)
3. Kato, S., Fujii, T., Koba, I., Sano, Y., Fu, K.I., Parra-Blanco, A., Tajiri, H., Yoshida, S., Rembacken, B.: Assessment of Colorectal Lesions Using Magnifying Colonoscopy and Mucosal Dye Spraying: Can Significant Lesions Be Distinguished? Endoscopy 33(3), 306–311 (2001)
4. Kato, S., Fu, K., Sano, Y., Fujii, T., Saito, Y., Matsuda, T., Koba, I., Yoshida, S., Fujimori, T.: Magnifying Colonoscopy as a Non-Biopsy Technique for Differential Diagnosis of Non-Neoplastic and Neoplastic Lesions. World of Gastroenterology 12(9), 1416–1420 (2006)

[2] This project was supported by the Austrian FWF Project, No.366-N15.

5. Hurlstone, D., Cross, S., Adam, I., Shorthouse, A., Brown, S., Sanders, D., Lobo, A.: Efficiacy of High Magnification Chromoscopic Colonoscopy for the Diagnosis of Neoplasia in Flat and Depressed Lesions of the Colorectum: a Prospective Analysis. Gut 53, 284–290 (2004)
6. Fu, K.: Chromoendoscopy using Indigo Carmine Dye Spraying with Magnifying Observation is the most Reliable Method for Differential Diagnosis between Non-Neoplastic and Neoplastic Colorectal Lesions. Endoscopy 36(12), 1089–1093 (2004)
7. Meining, A.: Inter- and Intra-Observer Variability of Magnification Chromoendoscopy for Detecting Specialized Intestinal Metaplasia at the Gastroesophageal Junction. Endoscopy 36(2), 160–164 (2004)
8. Kwitt, R., Uhl, A.: Modeling the marginal distributions of complex wavelet coefficient magnitudes for the classification of Zoom-Endoscopy images. In: Proceedings of the IEEE Computer Society Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA 2007), Rio de Janeiro, Brazil, pp. 1–8 (2007)
9. Kwitt, R., Uhl, A.: Color Eigen-Subband Features for Endoscopy Image Classification. In: Proceedings of the 33rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008), Las Vegas, Nevada, USA, pp. 589–592 (2008)
10. Van de Wouwer, G., Livens, S., Scheunders, P., Van Dyck, D.: Color Texture Classification by Wavelet Energy Correlation Signatures. In: Proceedings of the 9th International Conference on Image Analysis and Processing (ICIAP 1997), pp. 327–334. Springer, Heidelberg (1997)
11. Park, S.-H., Fürnkranz, J.: Efficient pairwise classification. In: Kok, J.N., Koronacki, J., Lopez de Mantaras, R., Matwin, S., Mladenič, D., Skowron, A. (eds.) ECML 2007. LNCS (LNAI), vol. 4701, pp. 658–665. Springer, Heidelberg (2007)
12. Fukunaga, K.: Introduction to Statistical Pattern Recognition. Morgan Kaufmann, San Francisco (1990)
13. Hansen, L., Salamon, P.: Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence 12(10), 993–1001 (1990)
14. Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)
15. Freund, Y., Schapire, R.: Experiments with a new boosting algorithm. In: Proceedings of the 13th International Conference on Machine Learning (ICML 1996), Bari, Italy, pp. 148–156 (1996)
16. Bay, S.: Nearest Neighbor Classification from Multiple Feature Subsets. Intelligent Data Analysis 3(3), 191–209 (1999)
17. Cutzu, F.: How to do multi-way classification with two-way classifiers. In: 3rd Joint International Conference on Artificial Neural Networks and Neural Information Processing, Istanbul, Turkey, pp. 375–384 (2003)
18. Dietterich, T.: Approximate statistical tests for comparing supervised classification learning algorithms. Neural Computation 10(7), 1895–1923 (1998)

Fast Automatic Segmentation of the Esophagus from 3D CT Data Using a Probabilistic Model

Johannes Feulner[1,3], S. Kevin Zhou[2], Alexander Cavallaro[4], Sascha Seifert[3], Joachim Hornegger[1,5], and Dorin Comaniciu[2]

[1] Chair of Pattern Recognition, University of Erlangen-Nuremberg, Germany
johannes.feulner@informatik.uni-erlangen.de
[2] Siemens Corporate Research, Princeton, NJ, USA
[3] Siemens Corporate Technology, Erlangen, Germany
[4] Radiology Institute, University Hospital Erlangen, Germany
[5] Erlangen Graduate School in Advanced Optical Technologies (SAOT), Germany

Abstract. Automated segmentation of the esophagus in CT images is of high value to radiologists for oncological examinations of the mediastinum. It can serve as a guideline and prevent confusion with pathological tissue. However, segmentation is a challenging problem due to low contrast and versatile appearance of the esophagus. In this paper, a two step method is proposed which first finds the approximate shape using a "detect and connect" approach. A classifier is trained to find short segments of the esophagus which are approximated by an elliptical model. Recently developed techniques in discriminative learning and pruning of the search space enable a rapid detection of possible esophagus segments. Prior shape knowledge of the complete esophagus is modeled using a Markov chain framework, which allows efficient inferrence of the approximate shape from the detected candidate segments. In a refinement step, the surface of the detected shape is non-rigidly deformed to better fit the organ boundaries. In contrast to previously proposed methods, no user interaction is required. It was evaluated on 117 datasets and achieves a mean segmentation error of 2.28mm with less than 9s computation time.

1 Introduction

The mediastinal region is of particular interest to radiologists for oncological examinations [1]. For diagnosis and therapy monitoring, CT scans of the thorax are common practice. Lymphoma, which is the second most common tumor in the mediastinum, often affects regions close to the trachea and the esophagus as these are natural gateways of the human body. While the trachea is very easy to see in CT, the esophagus is sometimes hard to find in single slices. Especially in coronal view, even experts have difficulties to see the boundaries, which is one reason why interpretation of the images is tedious. Fast and automatic segmentation of the esophagus can shorten the time a radiologist needs to read an image.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 255–262, 2009.

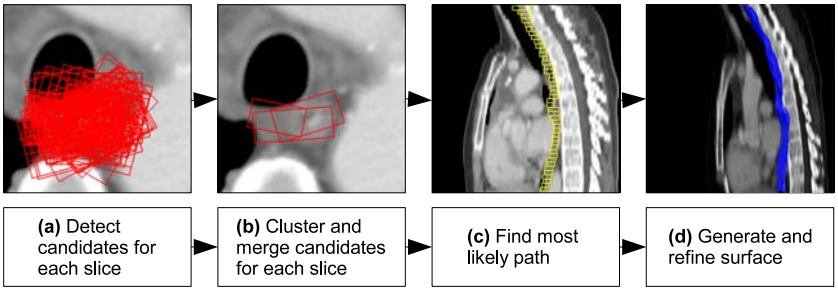


Fig. 1. Overview of the proposed method

Automated segmentation of the esophagus is challenging because of its complex shape and its inhomogeneous appearance. As its wall consists of muscle tissue, there is only little contrast if it is empty. Sometimes it is filled with air bubbles, remains of oral contrast agent or both. Up to now, there are few publications on esophagus segmentation, and all of them require a significant amount of user input. In [2], a probabilistic spatial and appearance model is used to extract the centerline. In a second step, the outer wall is approximated by fitting an ellipse into each slice using a region-based criterion. However, the method requires as input two points on the esophagus and furthermore a segmentation of the left atrium and the aorta. In [3], a semi-automated method is proposed that takes one contour in an axial slice as user input and propagates the contour to other slices by registration using optical flow. The quality of the segmentation was not evaluated quantitatively. Another semi-automated method is described in [4]. The user draws several contours in axial slices. The segmentation is obtained by interpolating the contours in the frequency domain. The image itself is not used.

In this work, a method is proposed that first detects the approximate shape. This is carried out in three sub steps, which are visualized in Figure 1 (a-c). In step (a), for each slice a detector that was trained from annotated data is run to detect weighted candidate esophagus segments, which are modeled as ellipses. An ellipse in visualized by its bounding box. These candidates are clustered to find modes in the distribution (b). Candidates of a cluster are merged into a weighted cluster center. A Markov chain model is then used to find the most likely path through the cluster centers. Prior shape knowledge is incorporated into the Markov chain by learning the transition probability distribution from a slice to the next from annotated data. In the final step (d), a surface is generated from the detected sequence of ellipses. A detector is trained offline to learn the boundary of the esophagus. The surface is deformed along its normals and smoothed to adapt the mesh to the organ boundary.

2 Esophagus Segmentation

2.1 Ellipse Detection

To approximate the contour of the esophagus, an elliptical model was chosen as it can be described by a relatively low dimensional parameter vector $\boldsymbol{e}$

$$\boldsymbol{e} = (x, y, \theta, a, b) \tag{1}$$

where x and y are the coordinates of the center within an axial slice, θ is the rotation angle and a and b are the semi-major and the semi-minor axis, respectively.

Recently developed techniques in discriminative learning [5] and search space pruning based on learning in marginal spaces [6] enable a rapid detection of candidate model instances. A probabilistic boosting tree (PBT) classifier was trained with a large number of positive and negative examples to learn the target distribution $p(m = 1|\boldsymbol{e}, \boldsymbol{v})$ which describes the probability that $\boldsymbol{e}$ is a correct model instance in the currently observed image $\boldsymbol{v}$. In order to accelerate search, a detector was also trained on the subspaces (x, y) and (x, y, θ) of the full parameter space $\boldsymbol{e}$ to learn the distributions $p(m = 1|(x, y))$ and $p(m = 1|(x, y, \theta))$. This allows to reject wrong model instances at an early stage. As feature pool, a combination of 3D Haar-like and steerable features were used [6]. Haar-like features are computed by convolving the image with box filter kernels of different size, position and weight. They gain their power from speed as they can be computed in constant time even for large kernels using an integral image. They are called Haar-like because of their similarity to the Haar wavelets. Steerable features are simple point features like intensity and gradient and nonlinear combinations of those evaluated at a certain sampling pattern, which is a regular grid of size $7 \times 7 \times 3$ in this case. The final output are the N best model instances $\boldsymbol{e}^{(i)}$, $i = 1 \ldots N$ together with a score $\varsigma^{(i)} = p(m = 1|\boldsymbol{e}^{(\boldsymbol{i})}, \boldsymbol{v})$ for each one.

In order to reduce subsequent search effort and to detect modes in the distribution of the candidates, they are clustered using an agglomerative hierarchical average-linkage clustering algorithm until a distance threshold $d_{\max}$ is reached, which was set to 10mm in the experiments. The distribution is now represented by the cluster centers $\boldsymbol{c}^{(1)} \ldots \boldsymbol{c}^{(K)}$ with weights $\sigma^{(1)} \ldots \sigma^{(K)}$, where the weight $\sigma^{(k)}$ of cluster center k is the sum of weights of all members.

2.2 Inferring the Path

So far, the axial slices of the volume image were treated separately. Shape knowledge is incorporated into a Markov chain model [7] of the esophagus, which is used to infer the most likely path through the axial slices. A graph of the Markov model used is depicted in Figure 2 (left). The variables $\boldsymbol{s}_1 \ldots \boldsymbol{s}_T$ correspond to the axial slices of the image. Possible states of a variable $\boldsymbol{s}_t$ are the ellipses corresponding to the cluster centers $\boldsymbol{c}_t^{(k)}, k = 1 \ldots K_t$ of slice t. Each state variable $\boldsymbol{s}_t$ is conditioned on the observed image slice $\boldsymbol{v}_t$. In Figure 2 (right), the factor graph [8] of the Markov model is shown. The clique potentials (or factors) of the

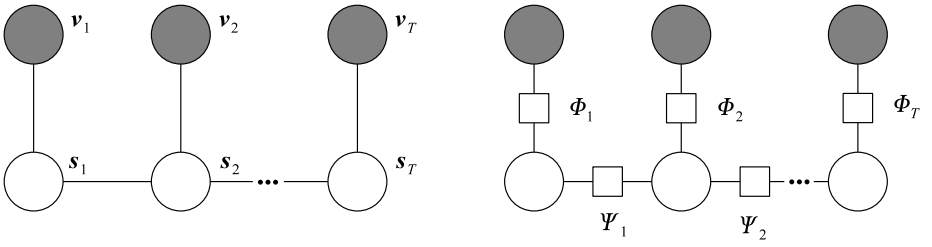


Fig. 2. Markov chain model of the esophagus along with corresponding factor graph

observation cliques are denoted with Φ_t. They are set to the score of the cluster centers:

$$\Phi_t(\boldsymbol{c}_t^{(k)}, \boldsymbol{v}_t) = \sigma_t^{(k)}. \tag{2}$$

The clique potentials Ψ_t between adjacent state variables $\boldsymbol{s}_t, \boldsymbol{s}_{t+1}$ represent the prior shape knowledge. They are set to the transition distribution from one slice to another:

$$\Psi_t(\boldsymbol{s}_t, \boldsymbol{s}_{t+1}) = p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t). \tag{3}$$

To simplify the transition distribution, it was assumed that the transition of the translation parameters x, y is statistically independent from the other parameters. The same was assumed for the scale parameters. As the rotation parameter θ is not well defined for approximately circular ellipses, the transition of rotation also depends on the scale parameters, but independence was assumed for rotation and translation parameters. With these assumptions, the transition distribution can be factorized and becomes

$$p(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t) = p(x_{t+1}, y_{t+1}|x_t, y_t)p(\theta_{t+1}|\theta_t, a_t, b_t)p(a_{t+1}, b_{t+1}|a_t, b_t). \tag{4}$$

The translation transition $p(x_{t+1}, y_{t+1}|x_t, y_t)$ is modeled as a 2D normal distribution $\mathcal{N}(\Delta x, \Delta y|\boldsymbol{\Sigma}_p, \boldsymbol{m}_p)$ and the scale transition $p(a_{t+1}, b_{t+1}|a_t, b_t)$ as a 4D normal distribution $\mathcal{N}(a_{t+1}, b_{t+1}, a_t, b_t|\boldsymbol{\Sigma}_s, \boldsymbol{m}_s)$. The variance of the rotation transition highly increases with the circularity of the ellipse as θ becomes arbitrary for a circle. Therefore, $p(\theta_{t+1}|\theta_t, a_t, b_t)$ is modeled with ten 1D normal distributions, one for a certain interval of circularity, which is measured by the ratio $\frac{b}{a}$:

$$p(\theta_{t+1}|\theta_t, a_t, b_t) = \mathcal{N}\left(\Delta\theta \middle| \sigma_r\left(\frac{b}{a}\right), m_r\left(\frac{b}{a}\right)\right). \tag{5}$$

The parameters of all normal distributions were estimated from manually annotated data.

The a posterior joint distribution of all states $p(\boldsymbol{s}_{1:T}|\boldsymbol{v}_{1:T})$ is then given by the product of all factors of the factor graph. The maximum a posteriori (MAP) estimate

$$\hat{\boldsymbol{s}}_{1:T}^{(\mathrm{MAP})} = \underset{\boldsymbol{s}_{1:T}}{\operatorname{argmax}} \left(\Phi_1(\boldsymbol{s}_1, \boldsymbol{v}_1) \prod_{t=2}^{T} \Phi_t(\boldsymbol{s}_t, \boldsymbol{v}_t)\Psi_{t-1}(\boldsymbol{s}_{t-1}, \boldsymbol{s}_t) \right) \tag{6}$$

can be computed efficiently using the max-sum algorithm, which is a variant of the sum-product algorithm [8].

2.3 Surface Generation and Refinement

After the MAP estimate of the path has been detected, the sequence of ellipses is converted into a triangular mesh representation by sampling the ellipses and connecting neighboring point sets with a triangle strip.

The cross-section of the esophagus is generally not an ellipse, and the path obtained in section 2.2 often contains some inaccuracies. Therefore, the mesh model is further refined to better fit the surface of the organ.

A PBT classifier was trained to learn the boundary of the esophagus. The classifier uses steerable features as proposed in [6]. As for ellipse detection, the steerable features are sampled on a regular grid, but now with a size of $5 \times 5 \times 9$. For each mesh vertex, the sampling pattern is placed so that the vertex is in the center of the pattern and the longest axis points in direction of the mesh normal. Now the pattern is moved along the normal to find the maximal detector response and the new position of the vertex. Finally, the surface is passed through a Laplacian smoothing filter. This process of deformation and smoothing is repeated for a certain number of iterations.

3 Results

The proposed method was evaluated on 117 CT scans of the thorax using threefold cross-validation. Manual segmentation was available for each dataset. The spatial resolution of the datasets was typically $0.72 \times 0.72 \times 5\text{mm}^3$. Among the scans, 34 were taken from patients suffering from lymphoma, which often causes enlarged lymph nodes in the mediastinal region. In some datasets, the esophagus contained remains of orally given contrast agent. For evaluation, the datasets were cropped around the region of interest.

Accuracy. The accuracy of the segmentation was evaluated by comparing the result with the annotated ground truth. Mean mesh-to-mesh distance and Hausdorff distance (maximal mesh-to-mesh distance) were used as error measures. Results are shown in Table 1. First, results are compared after the path inference step with surface refinement turned off (rows one to six). Accuracy was measured for $N = 100$ and $N = 200$ model instance candidates $\boldsymbol{e}^{(i)}$, $i = 1 \ldots N$. Additionally to a Markov model with a Markov order of one ($M = 1$), measurements for $M = 0$ and $M = 2$ are also included. As a 2nd order Markov chain over some alphabet is equivalent to a first order chain over the alphabet of 2-tuples, the model of Figure 2 was used also for the 2nd order case, but with a state space that consists of 2-tuples and with adapted transition probabilities.

While there is a noticeable improvement with $N = 200$ compared to $N = 100$, the Markov order has very little influence on the numerical results. However, the

Table 1. Accuracy of the registration in mm

Method	mean error	Hausdorff distance
$M = 0$, no refinement, $N = 100$	2.81 ± 1.19	15.5
$M = 1$, no refinement, $N = 100$	2.85 ± 1.31	15.6
$M = 2$, no refinement, $N = 100$	2.86 ± 1.30	15.8
$M = 0$, no refinement, $N = 200$	2.78 ± 1.12	15.2
$M = 1$, no refinement, $N = 200$	2.80 ± 1.30	15.1
$M = 2$, no refinement, $N = 200$	2.80 ± 1.26	15.0
$M = 1$, refinement, $N = 100$	2.35 ± 1.61	15.2
$M = 1$, refinement, $N = 200$	2.28 ± 1.58	14.5
Particle filtering, refinement, 1000 particles	4.84 ± 5.01	22.67

Table 2. The computation time in seconds is shown for the different steps of the algorithm with $M = 1$. N denotes the number of model candidates.

Method	Model detection	path inferrence	surface refinement	total
$N = 100$	3.67	0.0073	0.26	3.94
$N = 200$	7.99	0.0069	0.26	8.26

results generated with the Markov model turned on ($M = 1$ and $M = 2$) are visually more appealing because they are smooth and look more anatomically reasonable. As $M = 1$ and $M = 2$ produce very similar results, $M = 1$ is proposed as it does not introduce unnecessary complexity. The boundary refinement step significantly improves the segmentation error (rows seven and eight of Table 1). With $N = 200$ and $M = 1$, the proposed method gives a mean segmentation error of 2.28mm with a standard deviation of 1.58mm and a mean Hausdorff distance of 14.5mm. For comparison, the path of the esophagus was also detected using a particle filter approach [9] (last row of Table 1). Particle filtering is commonly used for tracking applications and is also becoming popular to detect tubular structures [10,11]. The Markov chain approach gives considerably better results because the image is searched exhaustively and thus it is far less prone to tracking loss.

Performance. Computation time was measured for the different steps of the algorithm. The results are summarized in Table 2. Time was measured on a 2.2GHz Intel Core2 Duo processor with 2GB of RAM on a volume of size $79 \times 96 \times 50$ voxels. Most of the time is spent on the model detection step, because here the volume is exhaustively searched. Computation time of this step increases linearly with the number N of candidates. In total, segmenting the esophagus takes 3.94s for $N = 100$ and 8.26s for $N = 200$.

Figure 3 (a-b) shows the segmentation error as a function of the number N of candidates and the number of surface refinement iterations. A value of $N = 200$ is a good trade-off between accuracy and speed, and two refinement iterations are a reasonable choice. Examples of segmentation results are displayed in Figure 3 (c-e).

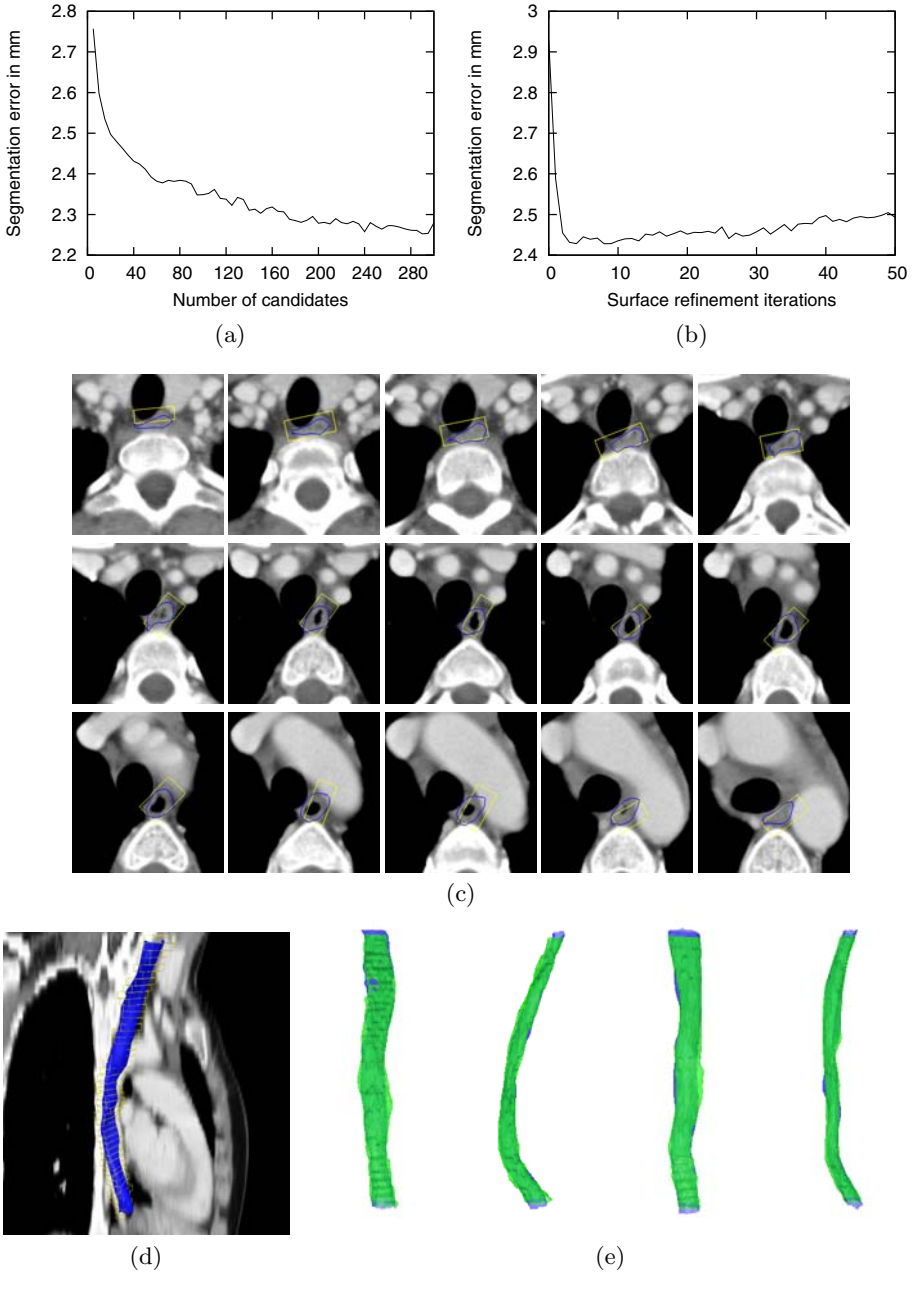


Fig. 3. (a-b): Segmentation error as a function of the number N of model candidates and the number of surface refinement iterations. (c-e): Examples of segmentation results. The boxes are bounding boxes of ellipses and visualize the inferred approximate shape. The final result after mesh generation and boundary deformation if shown in blue. In (e), the green semitransparent surface is the ground truth segmentation.

4 Discussion

The contribution of this work is twofold. First, the well known MAP framework for Markov chains is combined with the powerful detector based on the PBT classifier with Haar-like and steerable features [6]. Second, the method is extended with a boundary detector and applied to the problem of automatic esophagus segmentation, which is challenging due to the versatile shape and appearance of the organ.

With a mean segmentation error of 2.28mm, the proposed method has a good accuracy. Exhaustive search combined with a Markov model can well handle regions with clutter and low contrast. Compared to particle filtering based techniques, it is far less prone to tracking loss. Furthermore the method is fully automatic and very fast with a computation time of 8.3s. It can be easily adapted to other tubular structures like the spinal canal or larger vessels.

In the future we will consider to integrate more prior knowledge into the boundary refinement process. A local model seems most appropriate because otherwise cases where the esophagus is only partially visible cannot be easily handled any more.

References

1. Duwe, B.V., Sterman, D.H., Musani, A.I.: Tumors of the mediastinum. Chest 128(4), 2893–2909 (2005)
2. Rousson, M., Bai, Y., Xu, C., Sauer, F.: Probabilistic minimal path for automated esophagus segmentation. In: Proceedings of the SPIE, vol. 6144, pp. 1361–1369 (2006)
3. Huang, T.C., Zhang, G., Guerrero, T., Starkschall, G., Lin, K.P., Forster, K.: Semi-automated ct segmentation using optic flow and fourier interpolation techniques. Comput. Methods Prog. Biomed. 84(2-3), 124–134 (2006)
4. Fieselmann, A., Lautenschläger, S., Deinzer, F., John, M., Poppe, B.: Esophagus Segmentation by Spatially-Constrained Shape Interpolation. Bildverarbeitung für die Medizin, 247 (2008)
5. Tu, Z.: Probabilistic boosting-tree: learning discriminative models for classification, recognition, and clustering. In: ICCV, vol. 2, pp. 1589–1596 (2005)
6. Zheng, Y., Barbu, A., Georgescu, B., Scheuering, M., Comaniciu, D.: Fast automatic heart chamber segmentation from 3d ct data using marginal space learning and steerable features. In: ICCV, pp. 1–8 (2007)
7. Kindermann, R., Snell, J.L. (eds.): Markov Random Fields and Their Applications. AMS (1980)
8. Kschischang, F.R., Frey, B.J., Loeliger, H.-A.: Factor graphs and the sum-product algorithm. IEEE Transactions on Information Theory 47(2), 498–519 (2001)
9. Isard, M., Blake, A.: A smoothing filter for condensation. In: Burkhardt, H.-J., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1406, pp. 767–781. Springer, Heidelberg (1998)
10. Schaap, M., Smal, I., Metz, C., van Walsum, T., Niessen, W.: Bayesian tracking of elongated structures in 3d images. In: International Conference on Information Processing in Medical Imaging, IPMI (2007)
11. Florin, C., Paragios, N., Williams, J.: Particle filters, a quasi-monte-carlo-solution for segmentation of coronaries. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 246–253. Springer, Heidelberg (2005)

Automatic Segmentation of the Pulmonary Lobes from Fissures, Airways, and Lung Borders: Evaluation of Robustness against Missing Data

Eva M. van Rikxoort, Mathias Prokop, Bartjan de Hoop, Max A. Viergever, Josien P.W. Pluim, and Bram van Ginneken

Image Sciences Institute, University Medical Center Utrecht, The Netherlands

Abstract. Automatic segmentation of structures with missing or invisible borders is a challenging task. Since structures in the lungs are related, humans use contextual and shape information to infer the position of invisible borders. An example of a task in which the borders are often incomplete or invisible is the segmentation of the pulmonary lobes. In this paper, a fully automatic segmentation of the pulmonary lobes in chest CT scans is presented. The method is especially designed to be robust to incomplete fissures by incorporating contextual information from automatic lung, fissure, and bronchial tree segmentations, as well as shape information. Since the method relies on the result of automatic segmentations, it is important that the method is robust against failure of one or more of these segmentation methods. In an extensive experiment on 10 chest CT scans with manual segmentations, the robustness of the method to incomplete fissures and missing input segmentations is shown. In a second experiment on 100 chest CT scans with incomplete fissures, the method is shown to perform well.

1 Introduction

In medical images, structures are often incomplete, missing, or barely visible, so that they cannot be detected reliably by computer algorithms. In the lungs, anatomical structures are related; the position and shape of one anatomical structure provides information about the possible positions and shapes of other anatomical structures. Therefore, when the borders of structures in the lungs are not visible, human experts infer the position of the borders using anatomical cues in a flexible manner.

An example of a segmentation task for which the borders are often missing or hard to detect automatically is the segmentation of the pulmonary lobes. The human lungs are divided into five lobes, there are three lobes in the right lung and two in the left lung. The physical border between the lobes are the pulmonary fissures, the major fissures separate the upper and lower lobes in both lungs, the minor fissure in the right lung separates the middle lobe from the

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 263–271, 2009.

upper lobe. Several studies have shown that incomplete pulmonary fissures are a very frequent phenomenon [1]. In this work we present a method for automatic segmentation of the pulmonary lobes from computed tomography (CT) scans. Our method combines multiple anatomical cues to arrive at a segmentation, in an attempt to mimic the flexibility of human image understanding. A few papers concerning segmentation of pulmonary lobes have been published [2],[3],[4]. However, none of those papers present a completely automatic method for lobe segmentation in cases with incomplete fissures. In [4], a lobe segmentation in cases with incomplete fissures is presented that requires manual interaction in 25% of the cases.

Since fissures are often incomplete or hard to locate automatically, deriving a lobe segmentation from a fissure segmentation is not trivial. There are several alternatives that can be used to infer the lobe borders. Important information can be extracted from the segmentation of the lungs since the shape of the lobes is constrained by the shape of the lungs. In addition, the topology of the bronchial tree can be exploited. The different lobes are separately supplied by the first subdivisions of the bronchial tree after the main bronchi. Therefore, given a division of the bronchial tree into its lobar bronchi, a segmentation of the pulmonary lobes can be approximated by assigning each voxel inside the lungs to the lobar label of the nearest bronchus. Next to anatomical information, knowledge about the shape of the various lobes can be used. Radiologists combine all this information when they infer the lobar boundary in a CT scan.

We present a multi-atlas approach in which existing lobar segmentations are deformed to unseen test scans in which the fissures, as far as possible, the lungs, and the airway tree have been segmented. The key element of our method is the design of a cost function that exploits information from fissures, airway tree, and lung border in an effective way, such that less reliable information (lungs, airways) is only used when the most reliable information (fissures) is missing. To cope with the anatomical variation in the shapes of the lobes, an atlas selection mechanism is introduced.

The proposed method is fairly complex, in the sense that it requires the output of three segmentation algorithms as input. Each of these three tasks is not trivial by itself, therefore, it is likely that they will fail occasionally. Moreover, the most important information, derived from the detection of fissures, is often incomplete due to fissural incompleteness. Therefore, we carefully evaluate in this paper how robust the proposed method is to fissural incompleteness and missing lung, airway tree, and fissure segmentations. This also serves as a validation of the rationale of our approach: if these three structures *in concert* are indeed all important to infer the position of the lobar borders, leaving out any of this information should decrease the performance of the method. In an additional experiment on 100 chest CT scans, the performance of the method is evaluated.

2 Method

2.1 Overview

A multi-atlas lobe segmentation method was developed in which information from scans with complete fissures is transformed to a test scan to obtain an anatomically plausible lobe segmentation in the test scan. Since registering chest CT scans from different subjects directly does not lead to satisfactory results due to anatomical variations, anatomical information obtained from automatic segmentations of the lungs, the fissures, and the bronchial tree are employed during registration. The method requires a set of atlases with segmentations of the lung borders, fissures, and lobes to be available (Section 3). For a test image, the method consists of the following 5 steps: 1) the lung borders, fissures, and airway tree are segmented. 2) the most appropriate atlas is automatically selected. 3) the most appropriate atlas is registered to the test scan. 4) the lobe segmentation is propagated. 5) the result is post processed.

2.2 Prerequisite Segmentations

The method starts by segmenting the lung borders, fissures, and approximated lobe borders from the bronchial tree in the test image. The lung, fissure, and bronchial tree segmentations applied in this paper are all based on previous work ([5], [6], [7]) and are therefore not described here. From the segmentation of the lungs, the lung borders are extracted as those voxels in the lung segmentation for which one of the 8-connected neighbors is outside the lung segmentation. The segmentation of the bronchial tree is used to approximate the lobe borders. The method described by [7] provides the anatomical labels for airway branches up to segmental level. From these anatomical labels, the lobar label of each airway segment is know. Given this labeled airway, the lobes can be approximated by assigning each voxel in the lungs to the lobar label of the closest bronchi. Approximated lobar boundaries are extracted as those voxels for which one of the 8-connected neighbors in the lungs belongs to another lobe. The approximated lobe borders are illustrated in Figure 1.

Fig. 1. Illustration of the approximated lobe borders from the bronchial tree. The frames show the original slice, a rendering of the airways, the approximated lobes, and the approximated lobe borders, respectively.

2.3 Multi-atlas Lobe Segmentation

For the multi-atlas lobe segmentation it is assumed that a set of n atlases with complete pulmonary fissures and corresponding segmented lungs, fissures, and lobes are available (Section 3). For a test scan in which the lungs, fissures, and airways have been automatically segmented, the following steps are performed: 1) the test image and atlases are prepared for registration. 2) all atlases are registered to the test scan using a fast registration and the most promising atlas is selected for further processing. 3) the lobes in the test scan are segmented by accurately registering the selected atlas and propagating the lobe labels with the resulting transformation, followed by a postprocessing.

Registration. In this paper `elastix`[1] version 3.9 was used for registration. The atlas-based lobe segmentation requires two registration methods to be available: a fast (computationally cheap) method which is used to select the atlas that is most similar to the test image, and an accurate (computationally expensive) method which transforms an atlas image to a target image with high accuracy for the final lobe segmentation. For both the fast and the accurate registration SSD was used as a similarity measure. The images were first roughly aligned with an affine transformation, followed by a non-rigid registration modeled by B-splines. The difference between the fast and accurate registration is the number of iterations used per resolution for the non-rigid registration: five resolutions were used, the fast registration performed 100 iterations per resolution, the accurate registration performed 600 iterations per resolution.

Input to Registration. The registration is performed for two images simultaneously: one image containing the lung border and the other image containing the fissure segmentations. The cost function is calculated as a combination of both registrations, giving them both equal weight. Since registration involves an optimization towards a minimum of the similarity measure, not the binary segmentations are used as input but the distance transform to those segmentations. This smoothes the cost function and ensures that the method is gradually attracted to the correct position. The distance transforms to the different structures are cut off at 14 mm. For the test image, the fissure segmentations are augmented with the approximated lobe border from the bronchial tree only on positions where the fissures are not defined. This is achieved as follows: The distance transforms to the fissures and the approximated lobe borders are calculated separately. Next, the values of the distance transform to the approximated lobe border are added to the distance transform of the fissures only on those positions where the fissure distance transform is not defined.

Since SSD is used as a cost function, different structures can be given different weights during registration by multiplying their values by a weighing factor. For the lobe segmentation, three parameters are introduced controlling the weights of the different structures: ω_f for the fissure, ω_l for the lung border, and ω_b for the bronchial tree. The appropriate values for the different weights were found

[1] `elastix` is publicly available at `http://elastix.isi.uu.nl`.

to be $\omega_f = 4$, $\omega_l = 1$, and $\omega_b = 1$. This setting ensures that the fissures, when available, are the most important information used during the registration.

Atlas Selection. An atlas selection is performed in which the atlas that is anatomically most similar to the test scan is automatically selected. For this atlas selection, all atlases are registered to the test scan using fast registration, and the fissure segmentation of the atlases are transformed. The most promising atlas for the lobe segmentation of a particular test scan is the scan for which the fissures of the transformed scan line up best with the fissures of the test scan. The atlas with the highest number of voxels of the transformed fissure within 1 mm of the fissure in the test scan is selected to be the most promising atlas for further processing.

Lobe Segmentation. The lobes in the test scan are now segmented by registering the selected atlas to the test scan using accurate registration, and transforming the lobe segmentation available for the atlas to the test image. This results in a lobe segmentation for the test image. Since the lung borders of the atlas and test image are not forced to line up during the registration, a post processing is necessary. Every voxel inside the lung segmentation of the test image that is not assigned to a lobe after the transformation of the atlas lobes, is given the label of the closest voxel in the transformed lobe segmentation.

3 Materials and Experiments

Atlases. A set of five atlases with complete pulmonary fissures was taken from a lung cancer screening trial with low dose CT [8]. The lungs and fissures in the atlases were automatically segmented. The results of the automatic methods were visually checked and edited where needed by a human observer. Since the fissures in the test scans were complete, a lobe segmentation was extracted directly from the segmented fissures.

Test Data. Two test sets have been used in this paper, which we will refer to as test set A and test set B. Test set A contains 10 normal dose (120 kV, 100 to 150 mAs) inspiration CT chest scans of 10 different patients. The scans were selected to show (nearly) complete fissures in one of the lungs (visually confirmed). A human observer manually indicated the fissures in every fourth coronal slice. Test set B contains 100 low dose scans, selected by visual inspection to contain substantially incomplete fissures. All scans were reconstructed to 512×512 matrices, slice thickness 1 mm. To evaluate the performance and robustness of the automatic lobe segmentation three different experiments were performed:

1) To quantify the robustness of the automatic lobe segmentation to different levels of incomplete pulmonary fissures, an evaluation was performed on the scans from test set A. An experiment was setup in which parts of fissures were automatically removed to simulate the fissures being incomplete, for each lung, for each fissure separately. For a scan with complete fissures, the fissure was automatically segmented. Next, a part of the fissure was automatically removed from the hilum inward to mimic a scan with an incomplete fissure. Since the lobe

Table 1. Mean distance (mm) between the manually drawn lobe border (complete fissure) and the automatically found lobe border for different percentages of the fissure removed, for different settings of the weights ω_f, ω_l, and ω_b

ω_f-ω_l-ω_b	**setting**	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**	**70%**
		Left major fissure							
4-1-1	all structures	0.51	0.54	0.63	0.70	0.93	1.10	1.58	1.81
4-1-0	no airways	0.52	0.59	0.83	0.94	1.17	1.64	3.02	3.94
4-0-1	no lungs	0.48	0.56	0.67	0.73	0.99	1.27	1.68	1.99
0-1-1	no fissures	4.52	4.52	4.52	4.52	4.52	4.52	4.52	4.52
4-0-0	only fissures	0.49	0.57	0.84	1.06	1.31	1.89	3.73	5.47
0-1-0	only lungs	5.47	5.47	5.47	5.47	5.47	5.47	5.47	5.47
0-0-1	only airways	5.28	5.28	5.28	5.28	5.28	5.28	5.28	5.28
		Right major fissure							
4-1-1	all structures	1.22	1.32	1.54	1.79	2.18	2.56	3.22	3.71
4-1-0	no airways	1.25	1.39	1.70	2.13	2.83	3.76	5.28	7.31
4-0-1	no lungs	1.21	1.34	1.57	1.78	2.08	2.58	3.31	3.90
0-1-1	no fissures	6.33	6.33	6.33	6.33	6.33	6.33	6.33	6.33
4-0-0	only fissures	1.25	1.41	1.74	2.15	2.70	3.66	5.41	7.52
0-1-0	only lungs	6.62	6.62	6.62	6.62	6.62	6.62	6.62	6.62
0-0-1	only airways	6.31	6.31	6.31	6.31	6.31	6.31	6.31	6.31
		Right minor fissure							
4-1-1	all structures	1.08	1.02	1.37	2.00	2.60	3.38	4.61	5.90
4-1-0	no airways	1.08	1.07	1.61	2.30	3.21	4.15	6.56	8.06
4-0-1	no lungs	1.06	1.02	1.38	1.98	2.68	3.40	4.81	6.11
0-1-1	no fissures	6.23	6.23	6.23	6.23	6.23	6.23	6.23	6.23
4-0-0	only fissures	1.07	1.04	1.58	2.24	3.03	4.05	6.60	8.20
0-1-0	only lungs	6.93	6.93	6.93	6.93	6.93	6.93	6.93	6.93
0-0-1	only airways	6.23	6.23	6.23	6.23	6.23	6.23	6.23	6.23

segmentation method does not use the original CT data, it does not benefit from the fact that the fissure is actually complete. For the experiment, 10, 20, ... 70% of the fissures were removed. Both for the left and the right lung, six of the scans showed complete fissures and were used for this experiment. The mean distance between the manually defined fissures and the automatically determined lobe border was used as an evaluation measure.

2) To investigate the robustness of the method against missing lung, fissure, or airway segmentations, the first experiment was repeated using different settings for the weights ω_f, ω_l, and ω_b. The settings used are given in Table 1. For example, setting $\omega_f = 4$, $\omega_l = 1$, $\omega_b = 0$, does not take into account the approximated lobe border from the bronchial tree during the segmentation.

3) To evaluate the performance of the automatic lobe segmentation in a large set of scans, an observer study was setup in which a radiologist visually scored the performance of the lobe segmentation method for test set B. Each lobe border was scored to be correct, almost correct, or incorrect.

4 Results

In Table 1 the results for the first and second experiment are provided. Example output of the lobe segmentation for different percentages of fissures removed are provided in Figure 2. The results of the observer study in test set B were as follows: for the left major fissure, 79% of the lobe borders were scored as correct, the other 21% as almost correct. For the right major fissure and right minor fissure those numbers are 89%, and 11%, and 76%, and 22% respectively. For the right minor fissure two lobe borders were scored as incorrect. In both cases, the fissure was completely absent.

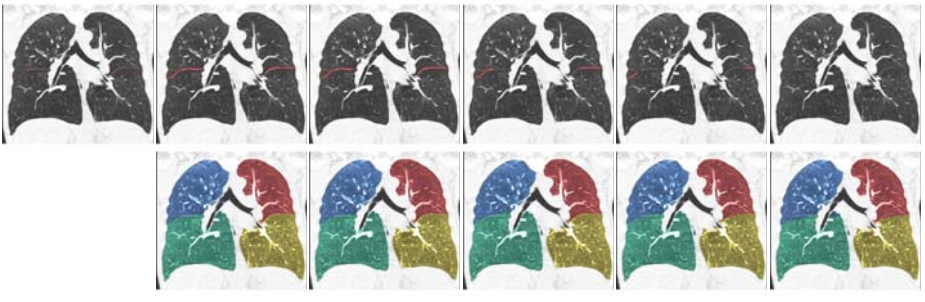


Fig. 2. Example output of the experiment in which parts of fissures were automatically removed to test the robustness of the automatic lobe segmentation against incomplete fissures. The first row shows the original image with the fissure segmentations, in the second row the resulting lobe segmentations are shown. The first column shows the ground truth. The next columns show the results for 0%, 10%, 30%, 50%, and 70% of the fissures removed, respectively.

5 Discussion and Conclusion

Complete fissures are rare, which makes automatic lobe segmentation challenging. In this paper, a completely automatic lobe segmentation method robust against the presence of incomplete pulmonary fissures was presented. By using an atlas based approach in which information from scans with complete fissures are transformed to test scans with incomplete fissures, the fissures are extended in an anatomically plausible manner. The methodology applied in this paper is similar to the methodology humans use when segmenting lobes manually: when the fissures are complete, they define the lobe borders. When no fissure is present, knowledge about the shapes of the various lobes as well as anatomical information, such as the bronchial tree, are used.

The results of the first two experiments clearly show that the presented automatic lobe segmentation performs well in cases with incomplete fissures. From Table 1 it can be seen that even when 70% of the fissures were removed, the automatically extracted lobe border was still within 4 mm of the real lobe border for the major fissures and within 6 mm for the right minor fissure, which

are clinically acceptable results. The accuracy of the automatic lobe segmentation gradually decreases when less fissures are present. In addition, the results shown in Figure 2 show that the lobe segmentation remains similar to the lobe segmentation with complete fissures even after removing 70% of the fissure.

The results in Table 1 also show the contribution of the different segmented structures used during registration. The first observation that can be made is that when the fissures are complete, they provide enough information to obtain an accurate lobe segmentation. Secondly, it can be seen that adding the approximated lobe border from the bronchial tree where fissures are incomplete leads to better performance: for all fissures, the performance decreases when ω_b is set to 0. In addition, when only the approximated lobe border from the bronchial tree is used, the average error is still within 7 mm for each fissure, indicating that it is a reasonable approximation of the position of the true lobe border. From Table 1 it can also be appreciated that the method is in general robust against missing input segmentations.

Overall, using the fissures, lungs, and bronchial tree as input gave the best results. Which confirm the rational behind our method that combining information from the fissures, lungs, and airways leads to superior results. This setting was used for a visual evaluation in the 100 scans from test set B. This evaluation showed that a radiologist agreed with the automatic result in 81% of the cases on average, only for two cases, only for the minor fissure, the segmentation was judged as incorrect. It is important to realize the difficulty of the task of automatic lobe segmentation with incomplete fissures. The scans in test set B were selected to show substantially incomplete fissures. As a comparison, Ukil et al [4] performed a similar evaluation for scans with incomplete fissures in which 40% of the lobe segmentations were judged as correct.

To conclude, an automatic lobe segmentation method was presented that was specifically designed to be robust against incomplete fissures by incorporating contextual information from automatic segmentations of other anatomical structures. Next to good performance in cases with incomplete fissures, the method was shown to be able to handle missing input segmentations.

References

1. Aziz, A., Ashizawa, K., Nagaoki, K., Hayashi, K.: High resolution CT anatomy of the pulmonary fissures. Journal of Thoracic Imaging 19(3), 186–191 (2004)
2. Kuhnigk, J.M., Dicken, V., Zidowitz, S., Bornemann, L., Kuemmerlen, B., Krass, S., Peitgen, H.O., Yuval, S., Jend, H.H., Rau, W.S., Achenbach, T.: New tools for computer assistance in thoracic CT part 1. Functional analysis of lungs, lung lobes and bronchopulmonary segments. Radiographics 25(2), 525–536 (2005)
3. Zhang, L., Hoffman, E.A., Reinhardt, J.M.: Atlas-driven lung lobe segmentation in volumetric x-ray CT images. IEEE Transactions on Medical Imaging 25(1), 1–16 (2006)
4. Ukil, S., Reinhardt, J.M.: Anatomy-guided lung lobe segmentation in x-ray CT images. IEEE Transactions on Medical Imaging 28(2), 202–214 (2009)

5. van Rikxoort, E.M., de Hoop, B., Viergever, M.A., Prokop, M., van Ginneken, B.: Automatic lung segmentation from thoracic ct scans using a hybrid approach with error detection. Medical Physics 36(7) (2009)
6. van Rikxoort, E.M., van Ginneken, B., Klik, M.A.J., Prokop, M.: Supervised enhancement filters: application to fissure detection in chest CT scans. IEEE Transactions on Medical Imaging 27(1), 1–10 (2008)
7. van Ginneken, B., Baggerman, W., van Rikxoort, E.M.: Robust segmentation and anatomical labeling of the airway tree from thoracic CT scans. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 219–226. Springer, Heidelberg (2008)
8. Xu, D.M., Gietema, H., de Koning, H., Vernhout, R., Nackaerts, K., Prokop, M., Weenink, C., Lammers, J., Groen, H., Oudkerk, M., van Klaveren, R.: Nodule management protocol of the NELSON randomised lung cancer screening trial. Lung Cancer 54(2), 177–184 (2006)

Joint Segmentation of Image Ensembles via Latent Atlases

Tammy Riklin Raviv[1], Koen Van Leemput[1,2,3], William M. Wells III[1,4], and Polina Golland[1]

[1] Computer Science and Artificial Intelligence Laboratory, MIT, USA
[2] Department of Information and Computer Science, Helsinki University of Technology, Finland
[3] Department of Neurology, MGH, Harvard Medical School, USA
[4] Brigham and Womens Hospital, Harvard Medical School, USA

Abstract. Spatial priors, such as probabilistic atlases, play an important role in MRI segmentation. However, the availability of comprehensive, reliable and suitable manual segmentations for atlas construction is limited. We therefore propose a joint segmentation of corresponding, aligned structures in the entire population that does not require a probability atlas. Instead, a latent atlas, initialized by a single manual segmentation, is inferred from the evolving segmentations of the ensemble. The proposed method is based on probabilistic principles but is solved using partial differential equations (PDEs) and energy minimization criteria. We evaluate the method by segmenting 50 brain MR volumes. Segmentation accuracy for cortical and subcortical structures approaches the quality of state-of-the-art atlas-based segmentation results, suggesting that the *latent atlas* method is a reasonable alternative when existing atlases are not compatible with the data to be processed.

1 Introduction

Probabilistic atlases are crucial for most MR segmentation methods due to the absence of well defined boundaries. Derived from comprehensive sets of manually labeled examples, atlases provide statistical priors for tissue classification and structure segmentation [1,9,13,14,18]. Although atlas-based segmentation methods often achieve accurate results, the need for spatial priors can be problematic. First, the availability of suitable atlases is limited since manual segmentation of a significant number of volumes requires expensive effort of an experienced physician. Second, the suitability of existing atlases for images from different populations is questionable. Examples include using normal adult brain atlas for brain parcellation of young children or patients with severe brain pathologies.

Recently, a few methods have been proposed to reduce or avoid the dependency on possibly incompatible atlases. In the atlas-based segmentation method in [3], topological constraints are used to avoid possible bias introduced by the atlas. In [19], manually labeled structures are used to support the automatic segmentation of neighboring structures within the same image. Tu *et al.* [17] propose a discriminative approach for the segmentation of adjacent brain structures using

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 272–280, 2009.

a set of discriminative features learned from training examples. Lord *et al.* [11] suggest a group-wise smoothing, segmentation and registration method for cross sectional MR scans. Bhatia *et al.* [5] update an initial atlas constructed from a different population using the evolving segmentations of multiple images.

Here we propose and demonstrate a method that does not use a set of training images or probabilistic atlases as priors. Instead we extract an ensemble of corresponding structures simultaneously. The evolving segmentation of the entire image set supports each of the individual segmentations. In practice, a subset of the model parameters, called the spatial parameters, is inferred as part of the joint segmentation processes. These latent spatial parameters, which can be viewed as a 'dynamic atlas', are estimated exclusively from the data at hand and a single manual segmentation. The latent atlas is used as a Markov Random Field (MRF) prior on the tissue labels. The main novelty of the method with respect to other group-wise segmentation methods such as [11,5] is the consistent statistically-driven variational framework for MR ensemble segmentation.

Our contribution is two-fold. We introduce a level set framework, that is based on probabilistic principles, in which segmentation uncertainty is expressed by the logistic function of the associated level set values, similar to [15]. We then use it for group-wise segmentation. We evaluate our method by segmenting the amygdala, temporal gyrus and hippocampus in each hemisphere in 50 MR brain scans. The dice scores achieved by our method approach the atlas-based segmentation results of [14].

2 Problem Definition and Probabilistic Model

Our objective is to segment a particular structure or region of interest in N aligned MR images. Specifically, we consider the 2-partition problem where each voxel in image I_n $(n = 1 \dots N)$ is assigned to either the foreground (structure of interest) or the background.

Let each image $I_n \colon \Omega \to \mathbb{R}^+$, be a gray level image with V voxels, defined on $\Omega \subset \mathbb{R}^3$ and $\Gamma_n \colon \Omega \to \{0, 1\}$ be the unknown segmentation of the image I_n. We assume that each Γ_n is generated iid from a probability distribution $p(\Gamma |\ \theta_\Gamma)$ where θ_Γ is a set of unknown parameters. We also assume that Γ_n generates the observed image I_n, independently of all other image-segmentation pairs, with probability $p(I_n|\Gamma_n, \theta_{I,n})$ where $\theta_{I,n}$ are the parameters corresponding to image I_n. We assign a specific set of intensity parameters to each image since the acquisition conditions might vary across subjects.

Let $\{I_1 \dots I_N\}$ be the given set of aligned images that form the observed variable in our problem and let $\Gamma = \{\Gamma_1 \dots \Gamma_N\}$ be the corresponding segmentations. The joint distribution $p(I_1 \dots I_N, \Gamma_1 \dots \Gamma_N|\Theta)$ is governed by the composite set of parameters $\Theta = \{\theta_\Gamma, \theta_{I,1} \dots \theta_{I,N}\}$. Our goal is to estimate the segmentations Γ.

We jointly optimize for the segmentations Γ and the parameters Θ, assuming that $I_1 \dots I_N$ are independent:

$$\{\hat{\Theta}, \hat{\Gamma}\} = \arg \max_{\{\Theta, \Gamma\}} \log p(I_1 \dots I_N, \Gamma_1 \dots \Gamma_N; \Theta) \tag{1}$$

$$= \arg \max_{\{\Theta,\Gamma\}} \sum_{n=1}^{N} \log p(I_n, \Gamma_n; \Theta) \tag{2}$$

$$= \arg \max_{\{\Theta,\Gamma\}} \sum_{n=1}^{N} \left[\log p(I_n|\ \Gamma_n; \theta_{I,n}) + \log p(\Gamma_n; \theta_\Gamma)\right]. \tag{3}$$

We propose to alternate between estimating the maximum a posteriori (MAP) segmentations and updating the model parameters. For a given value of the model parameters $\hat{\Theta}$, Equation (3) implies that the segmentations can be estimated by solving N separate MAP problems:

$$\hat{\Gamma}_n = \arg \max_{\Gamma_n} \left[\log p(I_n|\ \Gamma_n; \theta_{I,n}) + \log p(\Gamma_n;\ \theta_\Gamma)\right]. \tag{4}$$

We then fix $\hat{\Gamma}$ and estimate the model parameters $\Theta = \{\theta_\Gamma, \theta_{I,1}, \ldots \theta_{I,N}\}$ by solving two ML problems:

$$\hat{\theta}_{I,n} = \arg \max_{\theta_{I,n}} \log p(I_n|\ \Gamma_n, \theta_{I,n}) \tag{5}$$

$$\hat{\theta}_\Gamma = \arg \max_{\theta_\Gamma} \sum_{n=1}^{N} \log p(\Gamma_n|\ \theta_\Gamma). \tag{6}$$

In the following sections we present a level-set framework that is motivated by this probabilistic model. We reformulate the estimation problem stated in Eq. (4) such that the soft segmentations $p(\Gamma_n)$ rather then the Γ_n are estimated.

3 Probabilistic View of the Level Set Framework

Now we draw the connection between the probabilistic model presented above and the level set framework for segmentation. Let $\phi_n \colon \Omega \to \mathbb{R}$ denote a level set function associated with image I_n. The zero level $C_n = \{\mathbf{x} \in \Omega|\ \phi_n(\mathbf{x}) = 0\}$ defines the interface between the partitions of I_n. Vese and Chan [6] represent the image partitions by a regularized variant of the Heaviside function of ϕ_n, e.g., $\tilde{H}_\epsilon(\phi_n) = \frac{1}{2}(1 + \frac{2}{\pi} \arctan(\frac{\phi_n}{\epsilon}))$. Alternatively, we can use the hyperbolic tangent to achieve the same goal:

$$\tilde{H}_\epsilon(\phi_n) = \frac{1}{2}\left(1 + \tanh\left(\frac{\phi_n}{2\epsilon}\right)\right) = \frac{1}{1 + e^{-\phi_n/\epsilon}}. \tag{7}$$

For $\epsilon = 1$, the function $\tilde{H}_\epsilon(\cdot)$ is the logistic function. Similar to [15], we define the level set function ϕ_n using the log-odds formulation instead of the conventional signed distance function:

$$\phi_n(\mathbf{x}) \triangleq \epsilon\ \mathrm{logit}(p) = \epsilon\ \log \frac{p(\mathbf{x} \in w)}{1 - p(\mathbf{x} \in \omega)} = \epsilon\ \log \frac{p(\mathbf{x} \in \omega)}{p(\mathbf{x} \in \Omega \setminus \omega)}. \tag{8}$$

The scalar ϵ determines the scaling of the level set function ϕ_n with respect to the ratio of the probabilities. Substituting this definition into Eq. (7) we obtain

$$\tilde{H}_\epsilon(\phi_n(\mathbf{x})) = \frac{1}{1 + p(\mathbf{x} \in \Omega \setminus \omega)/p(\mathbf{x} \in \omega)} = p(\mathbf{x} \in \omega), \tag{9}$$

which implies that the function $\tilde{H}_\epsilon(\phi_n(\mathbf{x}))$ can be viewed as the probability that the voxel in location $\mathbf{x}$ belongs to the foreground region. The functions $\tilde{H}_\epsilon(\phi_n(\mathbf{x}))$ and $H(\phi_n(\mathbf{x}))$ therefore represent soft and hard segmentations, respectively. To simplify the notation we omit the subscript ϵ in the rest of the paper.

In the following subsections we relate the terms in Eq. (3) to the energy terms in the classical level set functional.

3.1 Image Likelihood Term

Let us first consider the image likelihood term in Eq. (3):

$$\log p(I_n|\ \Gamma_n, \hat{\theta}_{I,n}) = \sum_{\{v|\Gamma_n^v=1\}} \log p_{\text{in}}(I_n^v; \theta_{I,n}) + \sum_{\{v|\Gamma_n^v=0\}} \log p_{\text{out}}(I_n^v; \theta_{I,n}), \tag{10}$$

where p_{in} and p_{out} are the probability distributions of the foreground and background image intensities, respectively.

Let $E_I \cong -\log p(I_n|\ \Gamma_n, \hat{\theta}_{I,n})$ define the energy term associated with the image likelihood term. Using the level-set formulation and replacing the binary labels Γ_n in Eq. (10) with a soft segmentation represented by $\tilde{H}(\phi_n)$, we get:

$$E_I(\phi_n, \Theta) = -\int_\Omega \left[\log p_{\text{in}}(I_n; \theta_{I,n})\tilde{H}(\phi_n(\mathbf{x})) + \log p_{\text{out}}(I_n; \theta_{I,n})\tilde{H}\left(-\phi_n(\mathbf{x})\right)\right] d\mathbf{x}. \tag{11}$$

If we use, for example, Gaussian densities for p_{in} and p_{out} we get the familiar minimal variance term [6,12]. Here, we use a Gaussian mixture to model the background, as described later in the paper.

3.2 Spatial Prior Term

We define the prior probability $p(\Gamma_n|\ \theta_\Gamma)$ to be a Markov Random Field (MRF):

$$p(\Gamma_n|\theta_\Gamma) = \frac{1}{Z(\theta_\Gamma)} \prod_{v=1}^{V} (\theta_\Gamma^v)^{\Gamma_n^v} (1 - \theta_\Gamma^v)^{(1-\Gamma_n^v)} e^{-f(\Gamma_n^v, \Gamma_n^{\mathcal{N}(v)})}, \tag{12}$$

where $Z(\theta_\Gamma)$ is the partition function and $\mathcal{N}(v)$ are the closest neighbors of voxel v. The function $f(\cdot)$ accounts for the interactions between neighboring voxels. If we omit the pairwise term in Eq. (12), the prior on segmentations $p(\Gamma_n|\theta_\Gamma)$ reduces to a Bernoulli distribution, where the parameters θ_Γ represent the probability map for the structure of interest. The introduction of the pairwise clique potentials complicates the model but encourages smoother labeling configurations.

Using the continuous level set formulation with soft segmentation, we define the spatial energy term. as follows:

$$E_S(\phi_n, \Theta) = -\int_\Omega \left[\log \theta_\Gamma(\mathbf{x})\tilde{H}(\phi_n(\mathbf{x})) + \log(1-\theta_\Gamma(\mathbf{x}))\tilde{H}(-\phi_n(\mathbf{x}))\right] d\mathbf{x} \quad (13)$$

As in [4] we ignore the partition function, approximating the MRF model above. The logarithm of the pairwise clique potential term $f(\cdot)$ can be configured to act as a finite difference operator approximating the gradient of Γ_n at the voxel v [10]. It can be therefore viewed as an approximation of the continuous term

$$E_{\text{LEN}}(\phi_n) = \int_\Omega |\nabla \tilde{H}(\phi_n(\mathbf{x}))| d\mathbf{x}, \quad (14)$$

which is the commonly used smoothness constraint as reformulated in [6].

3.3 The Unified Energy Functional

We now construct the cost functional for $\phi_1 \dots \phi_N$ and the parameters Θ by combing Eqs. (11),(13) and (14):

$$E(\phi_1 \dots \phi_N, \Theta) = \gamma E_{\text{LEN}} + \beta E_I + \alpha E_S, \quad (15)$$

where $\alpha = 1 - \beta - \gamma$. As in [16] we tune the weights such that the contributions of the energy terms E_{LEN}, E_I and E_S to the overall cost are balanced.

4 Alternating Minimization Algorithm

We optimize the unified functional (15) in a set of alternating steps. For fixed model parameters Θ, the evolution of each of the level set functions ϕ_n follows the gradient descent equation:

$$\frac{\partial \phi_n}{\partial t} = \tilde{\delta}(\phi_n)\left\{\gamma \text{ div } (\frac{\nabla \phi_n}{|\nabla \phi_n|}) + \beta\left[\log p_{\text{in}}(I_n(\mathbf{x}); \hat{\theta}_{I,n}) - \log p_{\text{out}}(I_n(\mathbf{x}); \hat{\theta}_{I,n})\right]\right.$$
$$\left. + \alpha\left[\log \hat{\theta}_\Gamma - \log(1-\hat{\theta}_\Gamma)\right]\right\}, \quad (16)$$

where $\tilde{\delta}(\phi_n) \triangleq \tilde{\delta}_\epsilon(\phi_n) = \frac{d\tilde{H}_\epsilon(\phi_n)}{d\phi_n}$ is the derivative of the Heaviside function, i.e.,

$$\tilde{\delta}_\epsilon(\phi_n) = \frac{1}{2\epsilon}\text{sech}(\frac{\phi_n}{2\epsilon}) = \frac{1}{\epsilon \cosh(\frac{\phi_n}{\epsilon})}.$$

For fixed segmentations ϕ_n the model parameters are recovered by differentiating the cost functional (15) with respect to each parameter.

4.1 Intensity Parameters

We assume that the intensities of the structure of interest are drawn from a normal distribution, i.e., $p_{\text{in}}(I_n;\theta_{I,n}) = \mathcal{N}(I_n;\mu_n,\sigma_n^2)$. The distribution mean and variance of the foreground region of I_n are updated at every iteration. The intensities of the background tissues are modeled as a K-Gaussian mixture:

$$p_{\text{out}}(I_n;\theta_{I,n}) = \text{GMM}(\mu_n^1 \cdots \mu_n^K, \sigma_n^1 \cdots \sigma_n^K, w_n^1 \cdots w_n^K),$$

where w_n^k is the mixing proportion component k in the mixture. The Gaussian mixture model parameters are estimated using expectation maximization (EM) [7].

4.2 Spatial Parameters

We estimate the spatial function $\theta_\Gamma(\mathbf{x})$, constructing a dynamically evolving latent atlas, by optimizing the sum of the energy terms the depend on θ_Γ :

$$\hat{\theta}_\Gamma = \arg\max_{\theta_\Gamma} \sum_{n=1}^{N} \int_\Omega [\tilde{H}(\phi_n(\mathbf{x}))\log(\theta_\Gamma(\mathbf{x})) + (1-\tilde{H}(\phi_n(\mathbf{x})))\log(1-\theta_\Gamma(\mathbf{x}))]d\mathbf{x},$$

yielding $\hat{\theta}_\Gamma(\mathbf{x}) = \frac{1}{N}\sum_{n=1}^{N}\tilde{H}(\phi_n(\mathbf{x}))$.

5 Experimental Results

We test the proposed approach on 50 MR brain scans. Some of the subjects in this set are diagnosed with the first episode schizophrenia or affective disorder. The MR images (T1, $256 \times 256 \times 128$ volume, $0.9375 \times 0.9375 \times 1.5\text{mm}^3$ voxel size) were acquired by a 1.5-T General Electric Scanner. The data was originally acquired for brain morphometry study [8]. In addition to the MR volumes, manual segmentations of three structures (superior temporal gyrus, amygdala, and hippocampus) in each hemisphere were provided for each of the 50 individuals and used to evaluate the quality of the automatic segmentation results. MR images are preprocessed by skull stripping. The volumes were aligned using B-spline registration according to [2].

Assuming that the manual segmentation of a single instance is given, we initialize the latent atlas θ_Γ by the Heaviside function of a single manual segmentation smoothed with a Gaussian kernel of width σ ($\sigma = .35$ in our experiments). We ran the experiments twice, initializing the level-set functions of the signed distance function of the given manual segmentation or a sphere, with center and radius corresponding to the manually segmented structure. The results obtained using the first mode of initialization were slightly better. We excluded the image associated with the given manual segmentations from the ensemble. The algorithm was implemented in Matlab and ran on the average 7.5 minutes for 50 cropped volumes, excluding the time of the initial estimate of the background

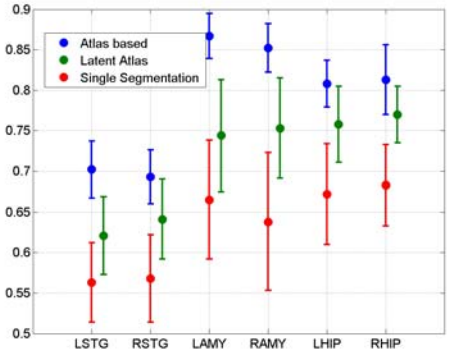


Fig. 1. The mean and standard deviation of the Dice scores calculated for six structures in the ensemble. The latent atlas segmentation (green) is compared to the atlas-based segmentation (blue) reported in [14] and to the segmentation obtained by using a single manual segmentation as an atlas (red).

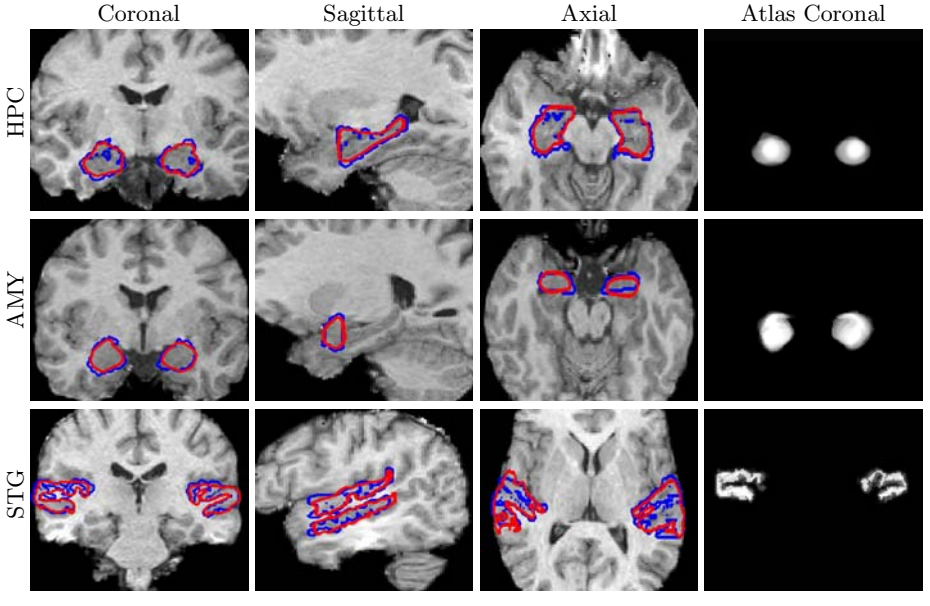


Fig. 2. Three cross-sections of 3D segmentations of Hippocampus, Amygdala and Superior Temporal Gyrus in the left and right hemispheres. Automatic segmentation is shown in red. Manual segmentation is shown in blue. Fourth column: Coronal views of the resulting atlases for each pair of structures.

intensity of the Gaussian mixture. About 7 iterations were needed until convergence which was obtained when the update of the level-set functions did not induce changes in the corresponding boundaries.

We used the Dice score to evaluate segmentations. The results are shown in Figure 1. In contrast to [14] we do not use spatial priors, neither do we use hierarchical multi-stage segmentation model. Nevertheless, the Dice scores obtained by our method approache the state-of-the-art atlas-based segmentation results reported there. To exemplify the significance of a latent atlas, generated concurrently with the segmentation, we also show a comparison to the segmentation results obtained by using an atlas constructed from a single manual segmentation smoothed by a Gaussian kernel, without the update procedure. Figure 2 shows segmentation examples of the three pairs of structures in representative individual brains. Coronal views of the resulting 3D atlases for each pair of structures are shown in the fourth column of Figure 2.

6 Discussion and Conclusions

We presented a level set framework for segmentation of MR image ensembles, that is motivated by a generative probabilistic model. Unlike most previous methods, we do not use spatial priors in the form of a probabilistic atlas. Instead, spatial latent parameters, which form a 'dynamic atlas', are inferred from the data set through an alternating minimization procedure.

The quality of the segmentation results obtained for ensembles of brain structures shows that the proposed method presents a reasonable alternative to standard segmentation techniques when a compatible atlas is not available.

An on-going research is now conducted to demonstrate the ability of the proposed algorithm to handle pathological cases (e.g., in a longitudinal or a multimodal study) where the atlas-based approach still fails.

Acknowledgments. This work was supported in part by NIH NIBIB NAMIC U54-EB005149, NIH NCRR NAC P41-RR13218, NIH NINDS R01-NS051826, NIH NCRR mBIRN U24-RR021382 grants and NSF CAREER Award 0642971.

References

1. Ashburner, J., Friston, K.: Unified segmentation. NeuroImage 26, 839–851 (2005)
2. Balci, S.K., Golland, P., Shenton, M., Wells, W.M.: Free-form B-spline deformation model for groupwise registration. In: Proc. of MICCAI 2007 Statistical Registration Workshop: Pair-wise and Group-wise Alignment and Atlas Formation, pp. 23–30 (2007)
3. Bazin, P.-L., Pham, D.L.: Statistical and topological atlas based brain image segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 94–101. Springer, Heidelberg (2007)
4. Besag, J.: Statistical analysis of non-lattice data. The Statistician 24(3), 179–195 (1975)
5. Bhatia, K.K., Aljabar, P., Boardman, J.P., Srinivasan, L., Murgasova, M., Counsell, S.J., Rutherford, M.A., Hajnal, J.V., Edwards, A.D., Rueckert, D.: Groupwise combined segmentation and registration for atlas construction. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 532–540. Springer, Heidelberg (2007)

6. Chan, T., Vese, L.: Active contours without edges. TIP 10(2), 266–277 (2001)
7. Dempster, A., Laird, N., Rubin, D.: Maximal likelihood form incomplete data via the EM algorithm. Proc. of the Royal Statistical Society 39, 1–38 (1977)
8. Hirayasu, Y., Shenton, M.E., Salisbury, D.F., Dickey, C.C., Fischer, I.A., Mazzoni, P., Kisler, T., Arakaki, H., Kwon, J.S., Anderson, J.E., Yurgelun-Todd, D., Tohen, M., McCarley, R.W.: Lower left temporal lobe MRI volumes in patients with first-episode schizophrenia compared with psychotic patients with first-episode affective disorder and normal subjects. Amer. J. Psychiatry 155(10), 1384–1391 (1998)
9. Fischl, B., Salat, D., Busa, E., Albert, M., Dietrich, M., Haselgrove, C., Van Der Kouwe, A., Killany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33(3), 341–355 (2002)
10. Li, S.: Markov Random Field Modeling in Computer Vision. Springer, Heidelberg (1995)
11. Lord, N., Ho, J., Vemuri, B.: Ussr: A unified framework for simultaneous smoothing, segmentation, and registration of multiple images. In: ICCV, pp. 1–6 (2007)
12. Paragios, N., Deriche, R.: Geodesic active regions: A new paradigm to deal with frame partition problems in computer vision. JVCIR 13, 249–268 (2002)
13. Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells, W.M.: A bayesian model for joint segmentation and registration. NeuroImage 31(1), 228–239 (2006)
14. Pohl, K.M., Bouix, S., Nakamura, M., Rohlfing, T., McCarley, R.W., Kikinis, R., Grimson, W.E.L., Shenton, M.E., Wells, W.M.: A hierarchical algorithm for MR brain image parcellation. TMI 26(9), 1201–1212 (2007)
15. Pohl, K.M., Fisher, J., Bouix, S., Shenton, M., McCarley, R., Grimson, W.E.L., Kikinis, R., Wells, W.M.: Using the logarithm of odds to define a vector space on probabilistic atlases. Medical Image Analysis 11(6), 465–477 (2007)
16. Riklin-Raviv, T., Sochen, N., Kiryati, N.: Shape-based mutual segmentation. International Journal of Computer Vision 79, 231–245 (2008)
17. Tu, Z., Narr, K.L., Dollár, P., Dinov, I., Thompson, P.M., Toga, A.W.: Brain anatomical structure segmentation by hybrid discriminative/generative models. IEEE TMI 27(4), 495–508 (2008)
18. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE TMI 18(10), 897–908 (1999)
19. Yang, J., Duncan, J.: Joint prior models of neighboring object for 3D image segmentation. In: CVPR, pp. 314–319 (2004)

Robust Medical Images Segmentation Using Learned Shape and Appearance Models

Ayman El-Baz[1] and Georgy Gimel'farb[2]

[1] Bioimaging Laboratory, Bioengineering Department, University of Louisville, Louisville, KY, USA
[2] Department of Computer Science, University of Auckland, Auckland, New Zealand

Abstract. We propose a novel parametric deformable model controlled by shape and visual appearance priors learned from a training subset of co-aligned medical images of goal objects. The shape prior is derived from a linear combination of vectors of distances between the training boundaries and their common centroid. The appearance prior considers gray levels within each training boundary as a sample of a Markov-Gibbs random field with pairwise interaction. Spatially homogeneous interaction geometry and Gibbs potentials are analytically estimated from the training data. To accurately separate a goal object from an arbitrary background, empirical marginal gray level distributions inside and outside of the boundary are modeled with adaptive linear combinations of discrete Gaussians (LCDG). Due to the analytical shape and appearance priors and a simple Expectation-Maximization procedure for getting the object and background LCDG, our segmentation is considerably faster than with most of the known geometric and parametric models. Experiments with various goal images confirm the robustness, accuracy, and speed of our approach.

1 Introduction

Parametric and geometric deformable models are widely used for image segmentation. However, in many applications, especially in medical image analysis, accurate segmentation with these models is a challenging problem due to noisy or low-contrast 2D/3D images with fuzzy boundaries between goal objects (e.g., anatomical structures) and their background; the similarly shaped objects with different visual appearances, and discontinuous boundaries because of occlusions or the similar visual appearance of adjacent parts of objects of different shapes [1,2]. Prior knowledge about the goal shape and/or visual appearance helps in solving such segmentation problems [2].

Relationship to the Prior Works. Initial attempts to involve the prior shape knowledge were built upon the edges. Pentland and Sclaroff [3] described an evolving curve with shape and pose parameters of a parametric set of points matched to strong image gradients and use a linear combination of eigenvectors to represent variations from an average shape. A parametric point model of Staib and Duncan [4] was based on an elliptic Fourier decomposition of landmarks.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 281–288, 2009.

Model parameters ensure the best match between the evolving curve and points of strong gradients.

More efficient results were obtained by learning the priors from a training set of manually segmented images of goal objects. Pizer et al. [5] and Styner et al. [6] segment 3D medical images by coarse-to-fine deformation of a shape-based medial representation ("m-rep"). A deformable model of Huang et al. [7] integrates region, shape and interior signal features assuming an approximate region shape is a priori known and aligned with the image to initialize the model. Leventon et al. [8] and Shen et al. [9] augment a level set-based energy function guiding the evolution with special terms attracting to more likely shapes specified with the principal component analysis (PCA) of the training set of goal objects, while Chen et al. [10] use a geometric model with the prior "average shape". The most advanced level set-based geometric model of Tsai et al. [1] evolves as zero level of a 2D map of the signed shortest distances between each pixel and the boundary. The goal shapes are approximated with a linear combination of the training distance maps for a set of mutually aligned training images. High dimensionality of the distance map space hinders PCA, and to simplify the model, only a few top-rank principal components are included to the linear combination [11].

Our approach follows the same ideas of using both the shape and appearance prior knowledge [1,13,12], but differs in the three aspects. **First**, instead of using the level set framework with linear combinations of the distance maps or statistical classifier (e.g., support vector machine), we use a simple parametric deformable model. Both the model and each goal shape are represented by piecewise-linear boundaries with a predefined number of control points. Corresponding points are positioned on roughly equiangular rays from the common center being the centroid of the control points along each boundary. A robust wave propagation is used to find correspondences in an aligned pair of the boundaries. **Secondly**, visual appearance of the goal objects is roughly described by characteristic statistics of gray level co–occurrences. Grayscale object pattern is considered as a sample of a spatially homogeneous Markov-Gibbs random field (MGRF) with multiple pairwise interaction. The interaction parameters of MGRF are estimated analytically. **Third**, the evolution is also guided at each step with a first-order probability model of the current appearance of a goal object and its background.

Basic Notation.

- (x, y) – Cartesian coordinates of
- $\mathbf{R} = [(x, y) : x = 0, \ldots, X-1; y = 0, \ldots, Y-1]$ – a finite arithmetic lattice supporting digital images and their region maps.
- $\mathbf{g} = [g_{x,y} : (x, y) \in \mathbf{R}; g_{x,y} \in \mathcal{Q}]$ – a grayscale digital image taking values from a finite set $\mathcal{Q} = \{0, \ldots, Q-1\}$.
- $\mathbf{m} = [m_{x,y} : (x, y) \in \mathbf{R}; m_{x,y} \in \mathbf{L}]$ – a region map taking labels from a binary set $\mathbf{L} = \{\mathsf{ob}, \mathsf{bg}\}$; each label $m_{x,y}$ indicates whether the pixel (x, y) in the corresponding image $\mathbf{g}$ belongs to a goal object (ob) or background (bg).

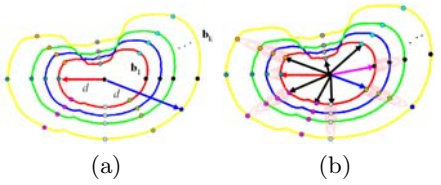
- $\mathbf{b} = [\mathbf{p}_k : k = 1, \ldots, K]$ – a deformable piecewise-linear boundary with the K control points $\mathbf{p}_k = (x_k, y_k)$ forming a circularly connected chain of line segments $(\mathbf{p}_1, \mathbf{p}_2), \ldots, (\mathbf{p}_{K-1}, \mathbf{p}_K), (\mathbf{p}_K, \mathbf{p}_1)$.
- $\mathbf{d} = [d_k^2 : k = 1, \ldots, K]$ – a vector description of the boundary $\mathbf{b}$ in terms of the square distances $d_k^2 = (x_k - x_0)^2 + (y_k - y_0)^2$ from the control points to the model centroid $\mathbf{p}_0 = (x_0 = \frac{1}{K}\sum_{k=1}^{K} x_k, y_0 = \frac{1}{K}\sum_{k=1}^{K} y_k)$, i.e. to the point at the minimum mean square distance from all the control points.
- $\mathbf{S} = \{(\mathbf{g}_t, \mathbf{m}_t, \mathbf{b}_t, \mathbf{d}_t) : t = 1, \ldots, T\}$ – a training set of grayscale images of the goal objects with manually prepared region maps and boundary models.
- $|\mathbf{A}|$ – the cardinality of a finite set $\mathbf{A}$.

2 Shape Prior

To build the shape prior, all the training objects in $\mathbf{S}$ are mutually aligned to have the same centroid and unified poses (orientations and scales of the objects boundaries) as in Fig. 1(a). For the definiteness, let each training boundary $\mathbf{b}_t \in \mathbf{S}$ is represented with K control points on the polar system of K° equiangular rays (i.e. with the angular pitch $2\pi/K^\circ$) emitted from the common centroid $\mathbf{p}_0$. The rays are enumerated clockwise, with zero angle for the first position $\mathbf{p}_{t,1}$ of each boundary. Generally, there may be rays with no or more than one intersection of a particular boundary, so that the number of the control points K may differ from the number of the rays K°. Because the training boundaries $\mathbf{b}_t \in \mathbf{S}$; $t = 1, \ldots, T$, share the same centroid $\mathbf{p}_0$, any linear combination $\mathbf{d} = \sum_{t=1}^{T} w_t \mathbf{d}_t$ of the training distance vectors defines a unique new boundary $\mathbf{b}$ with the same centroid. Typically, shapes of the training objects are very similar, and their linear combinations could be simplified by the PCA to escape singularities when adjusting to a given boundary.

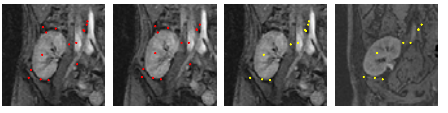
Let $\mathbf{D} = [\mathbf{d}_1\ \mathbf{d}_2\ \cdots \mathbf{d}_T]$ and $\mathbf{U} = \mathbf{D}\mathbf{D}^\mathsf{T}$ denote the $K \times T$ matrix with the training distance vectors as columns and the symmetric $K \times K$ Gram matrix of sums of squares and pair products $\sum_{t=1}^{T} d_{t,k} d_{t,k'}$; $k, k' = 1, \ldots, K$ of their components, respectively. The PCA of the matrix $\mathbf{U}$ produces K eigen-vectors $[\mathbf{e}_i : i = 1, \ldots, K]$ sorted by their eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_K \geq 0$. Due to identical or very similar training shapes, most of the bottom-rank eigenvalues are zero or very small, so that the corresponding "noise" eigenvectors can be discarded. Only a few top-rank eigenvectors actually represent the training shapes; the top distance eigenvector $\mathbf{e}_1$ corresponds to an "average" shape and a few others determine its basic variability. For simplicity, we select the top-rank subset of the eigenvectors $(\mathbf{e}_i : i = 1, \ldots, K')$; $K' < K$ by thresholding: $\sum_{i=1}^{K'} \lambda_i \approx \theta \sum_{i=1}^{K} \lambda_i$ with an empirical threshold $\theta = 0.8 \ldots 0.9$.

An arbitrary boundary $\mathbf{b}_\mathrm{c}$ aligned with the training set is described with the vector $\mathbf{d}_\mathrm{c}$ of the squared distances from its control points to the centroid. The prior shape approximating this boundary is specified by the linear combination of the training vectors: $\mathbf{d}^* = \sum_{i=1}^{K'} w_i^* \mathbf{e}_i \equiv \sum_{i=1}^{K'} \left(\mathbf{e}_i^\mathsf{T} \mathbf{d}_\mathrm{c}\right) \mathbf{e}_i$. Each signed difference



(a) (b)

Fig. 1. (a) Mutually aligned training boundaries and (b) searching for corresponding points between the mutually aligned training boundaries



(a–b) (a–c)

Fig. 2. Corresponding points found by SIFT in each pair (a,b), and (a,c) shown in red and yellow, respectively

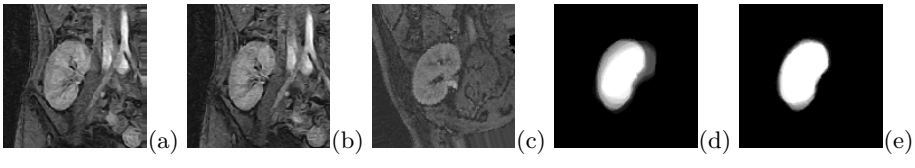
$\Delta_k = d_k^* - d_{c,k}$ determines the direction and force to move the boundary $\mathbf{b}_c$ towards the closest shape prior $\mathbf{b}^*$ specified by the distance vector $\mathbf{d}^*$.

Search for Corresponding Points is performed to suppress local "noise" (spurious deviations) in the training boundaries. The corresponding points are found in the aligned training boundaries by a robust wave-propagation based search (see Fig. 1(b)). An orthogonal wave is emitted from a point in one boundary, and the point at which the maximum curvature position of the wave front hits the second boundary is considered as the corresponding point.

SIFT-based Alignment. Just as the conventional level-set based geometric models with the shape priors, e.g. in [1], our approach depends essentially on accuracy of mutual alignment of similar shapes at both the training and segmentation stages. In the latter case the deformable model is initialized by aligning an image $\mathbf{g}$ to be segmented with one of the training images, say, $\mathbf{g}_1 \in \mathbf{S}$, arbitrarily chosen as a prototype.

First we use the *scale invariant feature transform* (SIFT) proposed by Lowe [14] to reliably determine a number of point-wise correspondences between two images under their relative affine geometric and local contrast / offset signal distortions. Then the affine transform aligning $\mathbf{g}$ most closely to $\mathbf{g}_1$ is determined by the gradient descent minimization of the mean squared positional error between the corresponding points.

Figure 2 shows correspondences found by SIFT in Dynamic Contrast Enhanced Resonance Imaging (DCE-MRI) images of human kidneys. The resulting affinely aligned goal shapes have roughly the same center and similar poses (orientations and scales). Quality of such alignment is evaluated in Fig. 3 by



(a) (b) (c) (d) (e)

Fig. 3. Mutually aligned images shown in Fig. 2 (a, b, c), and overlaps of the training region maps before (d) and after (e) the alignment

averaging all the training region maps $\mathbf{m}_t$; $t = 1, \ldots, T$, before and after the training set $\mathbf{S}$ is mutually aligned. It is clear that the alignment step reduced the variations between the co-aligned objects, SIFT based registration approach results in more accurate average shape model that is useful e.g. in any shape based segmentation framework.

3 Appearance Models

MGRF-based Appearance Prior. Our appearance prior is a rough descriptor of typically complex grayscale patterns of goal objects in terms of only second-order signal statistics. Each goal image is considered as a sample of a pairwise Markov–Gibbs random field (MGRF).

Let $\mathcal{N} = \{(\xi_i, \eta_i) : i = 1, \ldots, n\}$ be a finite set of (x, y)-offsets specifying neighbors $\{((x+\xi, y+\eta), (x-\xi, y-\eta)) : (\xi, \eta) \in \mathcal{N}\} \wedge \mathbf{R}$ interacting with each pixel $(x, y) \in \mathbf{R}$. Let $\mathcal{C}_{\xi,\eta}$ be a family of pairs $c_{\xi,\eta;x,y} = ((x, y), (x + \xi, y + \eta))$ in $\mathbf{R}$ with the offset $(\xi, \eta) \in \mathcal{N}$, i.e. the family of translation invariant pairwise cliques of the neighborhood graph on $\mathbf{R}$. Let $\mathbf{V}$ be a vector of Gibbs potentials for gray level co-occurrences in the neighboring pairs: $\mathbf{V}^{\mathsf{T}} = \left[\mathbf{V}_{\xi,\eta}^{\mathsf{T}} : (\xi, \eta) \in \mathcal{N}\right]$ where $\mathbf{V}_{\xi,\eta}^{\mathsf{T}} = \left[V_{\xi,\eta}(q, q') : (q, q') \in \mathcal{Q}^2\right]$. A generic MGRF with multiple pairwise interaction on $\mathbf{R}$ is specified by the Gibbs probability distribution (GPD)

$$P(\mathbf{g}) = \frac{1}{Z} \exp \sum_{(\xi,\eta)\in\mathcal{N}} \sum_{c_{\xi,\eta;x,y}\in\mathcal{C}_{\xi,\eta}} V_{\xi,\eta}(g_{x,y}, g_{x+\xi,y+\eta}) \equiv \frac{1}{Z} \exp |\mathbf{R}| \mathbf{V}^{\mathsf{T}} \mathbf{F}(\mathbf{g}) \quad (1)$$

Here, Z is the partition function, $\mathbf{F}^{\mathsf{T}}(\mathbf{g})$ is the vector of scaled empirical probability distributions of gray level co-occurrences over each clique family: $\mathbf{F}^{\mathsf{T}}(\mathbf{g}) = [\rho_{\xi,\eta}\mathbf{F}_{\xi,\eta}^{\mathsf{T}}(\mathbf{g}) : (\xi, \eta) \in \mathcal{N}]$ where $\mathbf{F}_{\xi,\eta}(\mathbf{g}) = [f_{\xi,\eta}(q, q'|\mathbf{g}) : (q, q') \in \mathcal{Q}^2]^{\mathsf{T}}$ and $\rho_{\xi,\eta} = \frac{|\mathcal{C}_{\xi,\eta}|}{|\mathbf{R}|}$ is the relative size of the clique family.

To identify the MGRF model described in Eq. (1), we have to estimate the Gibbs Potentials $\mathbf{V}$. In this paper we introduce a new analytical maximum likelihood estimation for the Gibbs potentials (the mathematical proof for this new estimator is shown on our website).

$$V_{\mathsf{ob};\xi,\eta}(q, q') = \lambda\big(f_{\mathsf{ob};\xi,\eta}(q, q') - 1/Q^2\big) \quad (2)$$

where λ is the analytically computed common scaling factor: $\lambda \approx Q^2$ if $Q \gg 1$ and $\rho_{\mathsf{ob};\xi,\eta} \approx 1$ for all $(\xi, \eta) \in \mathcal{N}$.

LCDG-models of Current Appearance. Non-linear intensity variations in a data acquisition system due to a scanner type and scanning parameters affect visual appearance of the object to be segmented. Thus, in addition to the learned appearance prior, we describe an on-going object appearance with a marginal intensity distribution within an evolving boundary $\mathbf{b}$ in $\mathbf{g}$. This distribution is considered as a dynamic mixture of two probability distributions that characterize the object and its background, respectively, and is partitioned into these two models using the EM-based approach in [15].

4 Model Evolution

The evolution $\mathbf{b}_\tau \rightarrow \mathbf{b}_{\tau+1}$ of the deformable boundary $\mathbf{b}$ in discrete time, $\tau = 0, 1, \ldots$, is determined by solving the Eikonal equation $|\nabla T(\mathbf{p}_{k,\tau})| F(\mathbf{p}_{k,\tau}) = 1$; $k = 1, \ldots, K$, where $F(\mathbf{p})$ is a speed function for the control point $\mathbf{p} = [x, y]$ of the current boundary. Our speed function depends on the shape prior, the LCDG-model of current appearance, and the MGRF-based appearance prior:

$$F(\mathbf{p}) = e^{-\beta|\Delta|} p_{\text{pix,ob}}(g_{x,y}) \pi_{x,y}(g_{x,y}|\mathbf{S}) \tag{3}$$

Here, Δ is the signed distance between the current control point $\mathbf{p} \in \mathbf{b}_\tau$ and the like one in the closest shape prior along the ray from the current boundary centroid. The constant factor β determines the evolution speed ($0 < \beta < 1$ for a smooth propagation). The marginal probabilities $p_{\text{pix,ob}}(q)$ and $p_{\text{pix,bg}}(q)$ of the gray value $g_{x,y} = q$ are estimated with the LCDG-submodels for the object and its background, respectively. The prior conditional probability $\pi_{x,y}(q|\mathbf{S})$ of the gray value $g_{x,y} = q$ in the pixel $\mathbf{p} = (x, y)$, given the current gray values in its neighborhood, is estimated in line with the MGRF prior appearance model:

$$\pi_{x,y}(g_{x,y}|\mathbf{S}) = \big(\exp\left(E_{x,y}(g_{x,y}|\mathbf{S})\right)\big) / \Big(\sum_{q \in \mathcal{Q}} \exp\left(E_{x,y}(q|\mathbf{S})\right)\Big)$$

where $E_{x,y}(q|\mathbf{S})$ is the pixel-wise Gibbs energy for the gray value q in the pixel $\mathbf{p} = (x, y)$, given the fixed gray values in its characteristic neighborhood: $E_{x,y}(q|\mathbf{S}) =$

$$\sum_{(\xi,\eta) \in \mathbf{N}'} \left(V_{\mathsf{ob};\xi,\eta}(g_{x-\xi,y-\eta}, q) + V_{\mathsf{ob};\xi,\eta}(q, g_{x+\xi,y+\eta})\right).$$

5 Experimental Results and Conclusions

Performance of the proposed parametric deformable model was evaluated on a large number of dynamic contrast-enhanced MRI (DCE-MRI) of human kidneys with the ground truth presented by a radiologist. The DCE-MRI are usually noisy, with continuously changing and low contrast. Note that about 20% of images were used to learn the prior models for the kidney. Basic segmentation stages of the algorithm are shown in Fig. 4. A comparative results in Table 1 for 4000 kidney DCE-MR images confirm accuracy and robustness of our approach. It is clear from Table 1 that the statistical analysis using a two tailed t-test shows that there is a significant difference ($P < 10^{-4}$) between the error generated by our segmentation approach and the error generated by the algorithm in [1] or ASM algorithm in [16]. Experiments with these DCE-MRI kidney images provide support for the proposed parametric deformable model guided with the learned shape and appearance priors. Our approach assumes that (*i*) the boundaries of the training and test objects are reasonably similar to within a relative affine transform and (*ii*) SIFT reliably detects corresponding points to automatically align the goal objects in the images despite their different backgrounds.

Table 1. Accuracy and time performance of our segmentation in comparison to Tsai et al.'s approach [1] and active shape model (ASM) [16]. One fifth of the kidney and corpus callosum images are used as training sets. Note that the training sets must contain all possible shape variability including pathology if it exists. Note that all these statistical results are calculated with respect to radiologist segmentation (ground truth).

	Seg. algorithm					
	Our	[1]	[16]	Our	[1]	[16]
Images(number)	Kidney (4000)			Corpus Callosum (200)		
Minimum error, %	**0.25**	3.9	1.3	**0.11**	1.9	0.9
Maximum error, %	**1.5**	8.3	10.6	**0.47**	4.1	4.9
Mean error, %	**0.83**	5.8	5.95	**0.27**	3.1	3.71
Standard deviation,%	**0.45**	1.5	3.7	**0.21**	0.97	1.17
Significant difference, P-value		$< 10^{-4}$	$< 10^{-4}$		$< 10^{-4}$	$< 10^{-4}$
Average time, sec	**23**	253	7	**11**	153	3

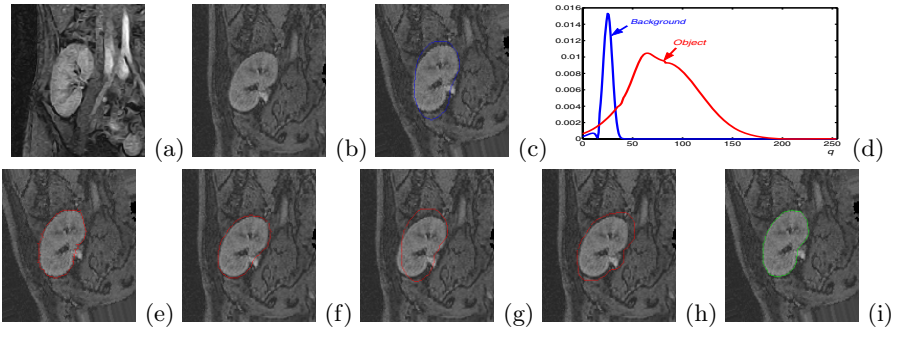


Fig. 4. Chosen training kidney prototype (a), an image to be segmented (b), its alignment to the prototype and initialization (in blue) (c), LCDG-estimates (d) of the marginal gray level distributions $\mathbf{p}_{\text{pix,ob}}$ and $\mathbf{p}_{\text{pix,bg}}$, final segmentation (in red) of the image aligned to the training prototype (e), the same result (f) after its inverse affine transform to the initial image (b) (the total error 0.63% comparing to the ground truth in green (i)), the segmentation (g) with the algorithm in [1] (the total error 4.9% comparing to the ground truth (i)), and the segmentation (h) with the algorithm in [16](the total error 5.17% comparing to the ground truth (radiologist segmentation) (i)). Note that our results obtained using 140 points describing kidney shape.

Although these assumptions restrict an application area of our approach comparing to the conventional parametric models, the latter typically fail on the above and similar images. More accurate level set-based geometric models with linear combinations of the training distance maps as the shape priors also rely on the mutual image alignment. Compared to these models, our approach escapes some of theoretical inconsistencies, is computationally much simpler and faster, and has the similar accuracy on high-contrast images, but notably better performance on low-contrast and multimodal ones. Due to space limitations, similar results for more complex medical images such as corpus callosum (see Table 1),

brain stem, and brain ventricles, are presented on our web site[1]. Also, we will post on our web site more results that demonstrate the effect of each model in the proposed segmentation framework. The proposed approach is suitable to segment pathological organs but a sample these pathological conditions must be included in the training data sets to be included in the prior shape model.

References

1. Tsai, A., Yezzi, A., Wells, W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A.: A shape based approach to the segmentation of medical imagery using level sets. IEEE TMI 22, 137–154 (2003)
2. Rousson, M., Paragios, N.: Shape priors for level set representations. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2351, pp. 78–92. Springer, Heidelberg (2002)
3. Pentland, A., Sclaroff, S.: Closed-form solutions for physically based shape modeling and recognition. IEEE PAMI 13, 715–729 (1991)
4. Staib, L., Duncan, J.: Boundary finding with parametrically deformable contour models. IEEE PAMI 14, 1061–1075 (1992)
5. Pizer, S., Gerig, G., Joshi, S., Aylward, S.: Multiscale medial shape-based analysis of image objects. Proc. IEEE 91, 1670–1679 (2003)
6. Styner, M., Gerig, G., Joshi, S., Pizer, S.: Automatic and robust computation of 3D medial models incorporating object variability. Int. J. Computer Vision 55, 107–122 (2002)
7. Huang, X., Metaxas, D., Chen, T.: MetaMorphs: Deformable shape and texture models. In: Proc. CVPR, vol. 1, pp. 496–503 (2004)
8. Leventon, M., Grimson, E., Faugeras, O.: Statistical shape influence in geodesic active contours. In: Proc. CVPR, pp. 316–323 (2000)
9. Shen, D., Davatzikos, C.: An adaptive-focus deformable model using statistical and geometric information. IEEE PAMI 22, 906–913 (2000)
10. Chen, Y., Thiruvenkadam, S., Tagare, H., Huang, F., Wilson, D., Geiser, E.: On the incorporation of shape priors into geometric active contours. In: IEEE Workshop Variational and Level Set Methods, pp. 145–152 (2001)
11. Pohl, K.M., Kikinis, R., Wells, W.M.: Active mean fields: Solving the mean field approximation in the level set framework. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 26–37. Springer, Heidelberg (2007)
12. Song, Z., Awate, S.P., Licht, D.J., Gee, J.C.: Clinical neonatal brain MRI segmentation using adaptive nonparametric data models and intensity-based markov priors. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 883–890. Springer, Heidelberg (2007)
13. Awate, S., Tasdizen, T., Foster, N., Whitaker, R.: Adaptive markov modeling for mutual-information-based, unsupervised MRI brain-tissue classification. Medical Image Analysis 10(5), 726–739 (2006)
14. Lowe, D.G.: Distinctive image features from scale-invariant Keypoints. Int. J. Computer Vision 60, 91–110 (2004)
15. El-Baz, A., Gimel'farb, G.: EM Based Approximation of Empirical Distributions with Linear Combinations of Discrete Gaussians. In: ICIP 2007, San Antonio, Texas, USA, September 16-19, 2007, vol. IV, pp. 373–376 (2007)
16. Cootes, T., Taylor, C.: A mixture model for representing shape variation. In: Proc. of British Machine Vision, pp. 110–119 (1997)

[1] http://louisville.edu/speed/bioengineering/faculty/bioengineering-full/dr-ayman-el-baz/dr-ayman-el-baz-publications.html

A Spatio-temporal Atlas of the Human Fetal Brain with Application to Tissue Segmentation

Piotr A. Habas[1,2], Kio Kim[1,2], Francois Rousseau[3],
Orit A. Glenn[2], A. James Barkovich[2], and Colin Studholme[1,2]

[1] Biomedical Image Computing Group
{piotr.habas,kio.kim,colin.studholme}@ucsf.edu
http://radiology.ucsf.edu/bicg
[2] Department of Radiology and Biomedical Imaging,
University of California San Francisco, San Francisco, CA 94143, USA
[3] LSIIT, UMR CNRS/ULP 7005, 67412 Illkirch, France

Abstract. Modeling and analysis of MR images of the early developing human brain is a challenge because of the transient nature of different tissue classes during brain growth. To address this issue, a statistical model that can capture the spatial variation of structures over time is needed. Here, we present an approach to building a spatio-temporal model of tissue distribution in the developing brain which can incorporate both developed tissues as well as transient tissue classes such as the germinal matrix by using constrained higher order polynomial models. This spatio-temporal model is created from a set of manual segmentations through groupwise registration and voxelwise non-linear modeling of tissue class membership, that allows us to represent the appearance as well as disappearance of the transient brain structures over time. Applying this model to atlas-based segmentation, we generate age-specific tissue probability maps and use them to initialize an EM segmentation of the fetal brain tissues. The approach is evaluated using clinical MR images of young fetuses with gestational ages ranging from 20.57 to 24.71 weeks. Results indicate improvement in performance of atlas-based EM segmentation provided by higher order temporal models that capture the variation of tissue occurrence over time.

1 Introduction

This work is motivated by the study of early human brain development in utero from fetal magnetic resonance (MR) imaging. Automated labeling of tissues is a key initial step in the accurate quantification of brain development. This allows us to model the normal development process and to extract rules to detect growth patterns that may be related to abnormal outcomes. In order to meaningfully label tissues present within a given brain image, it is necessary to interpret the anatomy in relation to its developmental stage. As identified by previous studies of neonates [1,2], infants [3], young children [4] and adults [5], the use of an age-specific atlas can significantly improve the results of automated analysis of brain image data.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 289–296, 2009.

The key challenges in segmentation of the early developing brain, as opposed to the developed adult brain, are the dramatic changes occurring over very short timescales [6]. These impose not simply shape changes but underlying changes in morphology whereby transient tissue classes appear and then disappear from different brain regions over time. Such cases would require the construction of many distinct atlases for different weeks of brain development.

Instead, here we describe an approach to modeling a complete 4-dimensional atlas of tissue distribution within the human fetal brain which can describe not only the variation in the presence of tissues, but also the appearance and complete disappearance of tissues classes over time through a temporally parametrized probability distribution for each voxel in the brain. Specifically, the approach makes use of a polynomial temporal models of tissue probabilities that can capture peaks in the occurrence of given tissue types over time.

In order to construct such a model, repeated serial imaging of the same anatomy is not a viable route. It is not feasible to repeatedly image pregnant women with a short enough time interval to capture subtle changes in morphology and morphometry. As a result, we have to use a large deformation groupwise diffeomorphic registration to bring collections of images of different fetuses with different ages into correspondence and construct models of common variation of tissue types over time.

2 Methods

2.1 MR Image Acquisition and Processing

Fetal MR imaging is performed in our institution on a 1.5T scanner without sedation or contrast agent administration. For each subject, multiple stacks of single-shot fast spin-echo (SSFSE) T2-weighted slice images (in plane resolution $0.469\text{mm} \times 0.469\text{mm}$, slice thickness $\approx 3\text{mm}$) are acquired in the approximately axial, sagittal and coronal plane with respect to the fetal brain. The MR sequence parameters ($\text{TR} = 4000 - 8000\text{ms}$, $\text{TE} = 91\text{ms}$) were originally designed for clinical scans and cannot be adjusted for image analysis purposes. After acquisition, image stacks are registered using an intersection-based technique [7] to account for spontaneous fetal movement during scanning and reconstructed into 3D volumes with isotropic resolution $0.469\text{mm} \times 0.469\text{mm} \times 0.469\text{mm}$. Reconstructed volumes are manually segmented into regions of cortical grey matter (GM), white matter (WM), the germinal matrix (GMAT) and ventricles (VENT).

2.2 Groupwise Registration of Segmentations

In order to relate data from multiple fetal subjects with different anatomies and tissue morphologies, a groupwise registration is performed between the tissue segmentations of the training dataset. Each base WM map is first globally aligned to the WM map of one of the subjects using an affine transformation of each anatomy driven by maximization of the normalized mutual information [8]. This provides a starting estimate for further collective alignment of the subjects

into a common average space. In order to avoid extreme warping between subjects with and without the germinal matrix in different brain regions, only two basic tissue classes are used for this registration – grey matter (GM), and white matter/germinal matrix combined (WM+GMAT). Each binary tissue region in each subject s is blurred and used to form a smooth map of tissue probability $\mathbf{p}_s(\mathbf{x}) = \left[p_{\mathrm{GM},s}(\mathbf{x})\ p_{\mathrm{WM+GMAT},s}(\mathbf{x}) \right]$ with components varying between 0 and 1 and representing the presence or absence of GM and WM+GMAT, respectively.

A combined registration metric

$$\mathcal{D}_s(\mathbf{x}) = -|\mathbf{p}_s(\mathbf{x} + \mathbf{u}_s(\mathbf{x})) - \bar{\mathbf{p}}(\mathbf{x})|^2 \tag{1}$$

describing the similarity between each set of subject tissue maps $\mathbf{p}_s(\mathbf{x})$ and the current group-averaged tissue map $\bar{\mathbf{p}}(\mathbf{x})$ is formed from the sum of squared differences in probability of the two tissue classes in each pair of subjects. The derivative of (1) with respect to the spatial mapping $\mathbf{u}_s(\mathbf{x})$ between subject maps is used as a force to drive a symmetric groupwise multiple-elastic registration algorithm. The tissue content difference force is balanced by a linear elastic energy

$$\mu\nabla^2\mathbf{u}_s(\mathbf{x}) + (\lambda + \mu)\nabla(\nabla^T\mathbf{u}_s(\mathbf{x})) + \beta\mathcal{S}_\mathbf{s}(\mathbf{x}) = \nabla_\mathbf{u}\mathcal{D}_\mathbf{s}(\mathbf{x}) \tag{2}$$

where μ and λ are constants determining the strength of the regularization. The resulting set of diffemorphic maps provides mapping between the average shape of the group and each of the training anatomies.

2.3 Voxelwise Statistical Modeling

After spatial normalization, we perform statistical modeling to create a probabilistic spatio-temporal model of tissue distribution in the fetal brain. Based on a set of s segmented images corresponding to gestational ages t_s, the model uses the age of the fetus as an independent variable and tissue class probabilities $p_{ks}(\mathbf{x})$ derived from manual segmentations as target values for each voxel $\mathbf{x}$ in the average shape space. As statistical modeling directly in probability space may result in values that are not valid probabilities as shown in Fig. 1, we use an alternative LogOdds representation of probabilities [9].

The LogOdds $\mathcal{L}(p_{ks}(\mathbf{x}))$ of a probability $p_{ks}(\mathbf{x})$ is the logarithm of the odds between the probability $p_{ks}(\mathbf{x})$ and its complement $\bar{p}_{ks}(\mathbf{x}) = 1 - p_{ks}(\mathbf{x})$.

$$\mathcal{L}(p_{ks}(\mathbf{x})) = \mathrm{logit}(p_{ks}(\mathbf{x})) = \log\left(\frac{p_{ks}(\mathbf{x})}{1 - p_{ks}(\mathbf{x})}\right) = l_{ks}(\mathbf{x}) \tag{3}$$

In the context of brain segmentation, $p_{ks}(\mathbf{x})$ is the probability that voxel $\mathbf{x}$ in subject s is assigned to tissue class k whereas its complement $\bar{p}_{ks}(\mathbf{x})$ is the probability of the assignment to any other tissue type. The inverse of the LogOdds function $\mathcal{L}(\cdot)$ is the standard logistic function

$$\mathcal{L}^{-1}(l_{ks}(\mathbf{x})) = \mathrm{logsig}(l_{ks}(\mathbf{x})) = \frac{1}{1 + \exp(-l_{ks}(\mathbf{x}))} = p_{ks}(\mathbf{x}) \tag{4}$$

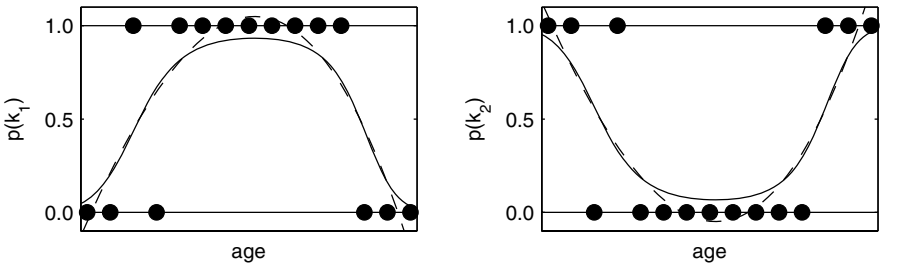


Fig. 1. An example of fitting a 3rd degree polynomial to probabilities of two mutually exclusive classes k_1 and k_2 (dots) directly in probability space (dashed line, note invalid probability values $p < 0$ and $p > 1$) and in the space of LogOdds (solid line)

A spatio-temporal atlas is created by temporal modeling of changes in tissue probabilities $p_{ks}(\mathbf{x})$ at each point $\mathbf{x}$ of the average shape space. For each $\mathbf{x}$, tissue class probabilities $p_{ks}(\mathbf{x})$ ($k \in \{1, 2, \ldots, K\}$) are transfered first to the space of LogOdds using (3). Then, a temporal model $\hat{l_{ks}}(\mathbf{x}, t)$ is obtained by interpolation between values of $l_{ks}(\mathbf{x})$ corresponding to gestational ages t_s of the s subjects. Time-related changes in probabilities $p_k(\mathbf{x})$ are modeled using a D-th degree polynomial in the space of LogOdds

$$\hat{l_k}(\mathbf{x}, t) = \sum_{d=0}^{D} a_{kd}(\mathbf{x}) t^d \tag{5}$$

with coefficients $a_{kd}(\mathbf{x})$ found through least squares fitting. To obtain a full probabilistic spatio-temporal model $\hat{\mathbf{p}}(\mathbf{x}, t) = \left[\hat{p_1}(\mathbf{x}, t)\ \hat{p_2}(\mathbf{x}, t) \ldots \hat{p_K}(\mathbf{x}, t)\right]$, all $\hat{l_k}(\mathbf{x}, t)$ need to be transfered back to probability space using (4) and normalized across K tissue classes to assure that $\hat{p_k}(\mathbf{x}, t)$ are legitimate probabilities.

$$\hat{p_k}(\mathbf{x}, t) = \frac{\mathcal{L}^{-1}(\hat{l_k}(\mathbf{x}, t))}{\sum_k \mathcal{L}^{-1}(\hat{l_k}(\mathbf{x}, t))} \tag{6}$$

The spatio-temporal model $\hat{\mathbf{p}}(\mathbf{x}, t)$ can be used for generation of age-specific probabilistic atlases $\hat{\mathbf{p}}(\mathbf{x})$ for any value of gestational age t. Such tissue probability maps may serve as sources of spatially varying priors for atlas-based segmentation of the fetal brain.

3 Results

3.1 Population

Clinical MR imaging was performed on a 1.5T scanner for a group of 10 fetal subjects with normal brain development. The gestational ages of the fetuses were almost uniformly distributed between 20.57 and 24.71 weeks (Fig. 2). For each

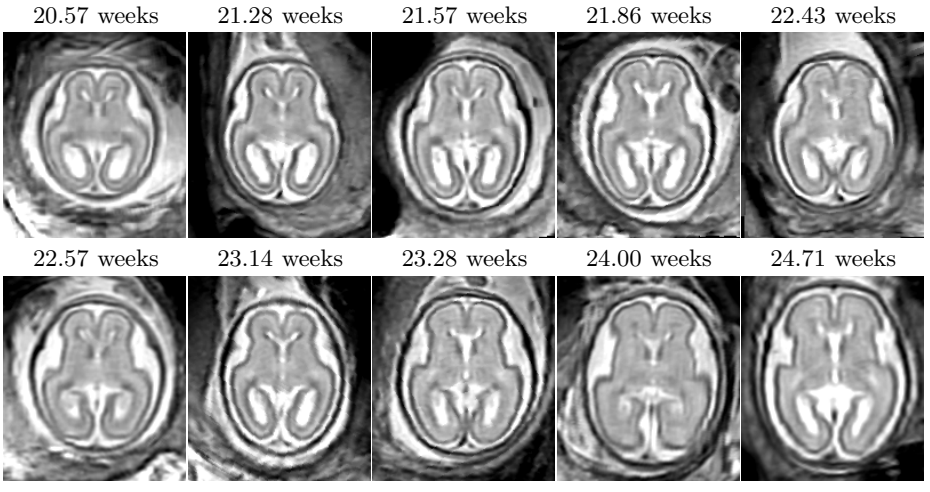


Fig. 2. Axial views of rigidly aligned reconstructed MR images of 10 fetal subjects ($20.57 - 24.71$ weeks GA) demonstrating variability in brain size and shape

subject, a high-resolution volume was reconstructed and manually segmented into regions of grey matter (GM), white matter (WM), the germinal matrix (GMAT) and ventricles (VENT). The resulting tissue label maps were verified by pediatric neuroradiologists with experience in fetal imaging and served as reference for further experiments. The reconstructed volumes were spatially normalized by groupwise registration of their respective manual segmentations as described in Section 2.2. After registration, the tissue label maps were transformed to the average space using nearest-neighbor interpolation.

3.2 Model Construction and Generation of Age-Specific Atlases

Based on the values of $p_{ks}(\mathbf{x})$ and the corresponding gestational ages t_s of the 10 fetal subjects, we created a spatio-temporal model $\hat{\mathbf{p}}(\mathbf{x}, t)$ with $D = 0$. This is equivalent to simple averaging of spatially normalized tissue label maps as is conventionally performed in atlas construction. The spatially varying tissue probability maps generated from such a model are not time-dependent and are shown in Fig. 3. To model dynamic changes in tissue presence at different gestational ages, we also created spatio-temporal models $\hat{\mathbf{p}}(\mathbf{x}, t)$ with $D = 1$ and $D = 2$ (higher order models with $D > 2$ were unstable for the available number of training subjects). From these models, we generated sets of age-specific probabilistic atlases $\hat{\mathbf{p}}(\mathbf{x})$, examples of which are shown in Fig. 4. Visual inspection of these tissue probability maps confirms that higher order spatio-temporal models correctly capture age-related changes in the morphology of the fetal brain such as the rapidly decreasing volume of the germinal matrix (especially in the areas of the occipital lobes) and shrinking of ventricles with respect to the overall brain volume.

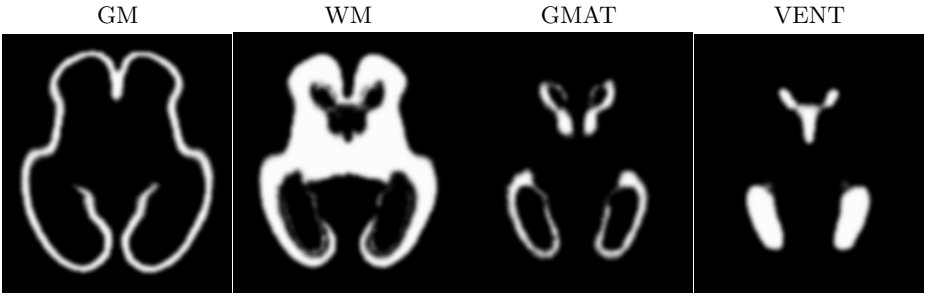


Fig. 3. Tissue probability maps generated from a spatio-temporal model with $D = 0$. This is equivalent to simple averaging of spatially normalized tissue label maps.

3.3 Application to Atlas-Based Tissue Segmentation

Based on previously reported results for infants [3] and young children [4], we expected that the use of age-specific atlases may also improve the performance of atlas-based segmentation of the fetal brain. To verify this hypothesis, we transformed MR images of the 10 fetal subjects to the average space and automatically segmented them using conventional EM algorithm [10,11]. For each subject, we generated age-matched probabilistic atlases $\hat{\mathbf{p}}(\mathbf{x})$ from spatio-temporal models $\hat{\mathbf{p}}(\mathbf{x}, t)$ with $D = 0$, $D = 1$ and $D = 2$ and used them as sources of spatially-varying tissue priors to drive the segmentation process.

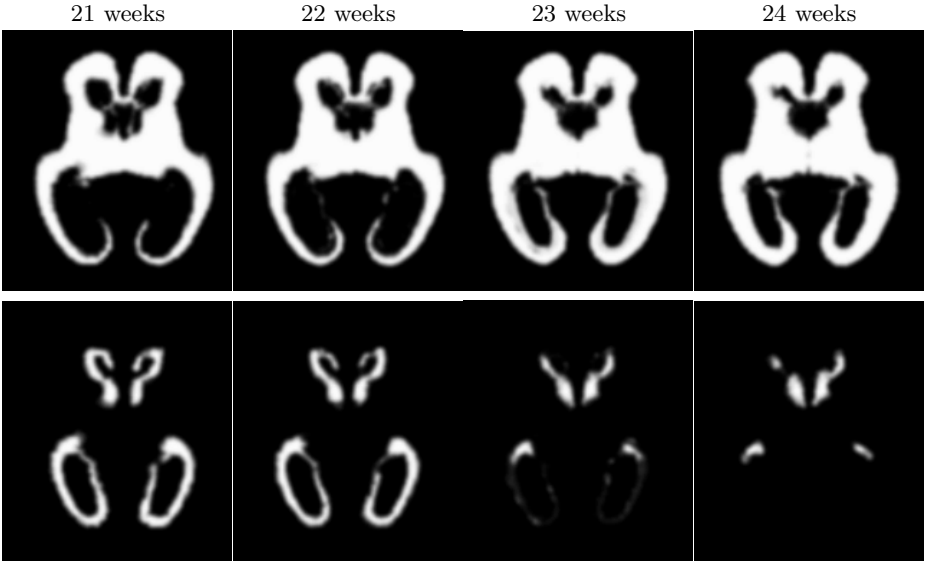


Fig. 4. Age-specific tissue probability maps for white matter (top row) and the germinal matrix (bottom row) generated from a spatio-temporal model with $D = 2$

Table 1. Average DSC performance of atlas-based EM segmentation of 10 fetal subjects with probabilistic atlases generated from spatio-temporal models of degree D. The p values represent the significance of improvement vs. simple averaging ($D = 0$).

Structure	D = 0	D = 1		D = 2	
	DSC	DSC	p	DSC	p
GM	0.834 ± 0.014	0.835 ± 0.013	0.168	0.836 ± 0.013	0.180
GMAT	0.642 ± 0.131	0.752 ± 0.069	**0.012**	0.761 ± 0.081	**0.011**
WM	0.905 ± 0.012	0.917 ± 0.014	**0.038**	0.918 ± 0.014	**0.034**
VENT	0.857 ± 0.040	0.862 ± 0.029	0.411	0.867 ± 0.029	0.099

For quantitative evaluation, we calculated the values of the Dice similarity coefficient (DSC) [12] between the results of automatic segmentation and respective manual segmentations of each subject. Table 1 presents the average DSC performance of automatic EM segmentation with probabilistic atlases generated from spatio-temporal models with $D = 0$, $D = 1$ and $D = 2$ for four different types of fetal brain structures. For $D = 1$ and $D = 2$, the significance of DSC improvements with respect to $D = 0$ was determined by a paired t-test.

For all degrees of the spatio-temporal model, automatic atlas-based EM segmentation yields acceptable performance (DSC > 0.8) for grey matter, white matter and ventricular CSF. Similar DSC values were previously reported for automatic brain segmentation in neonates [1,2] and infants [3]. The inferior average performance for the germinal matrix (DSC $= 0.642$) for $D = 0$ arises mainly from oversegmenting of this tissue type for subjects older than 23 weeks. For these subjects, up to 8% of voxels actually representing white matter is incorrectly assigned to GMAT, bringing down the average performance for this subgroup (DSC $= 0.475$).

The use of age-specific atlases generated from spatio-temporal models with $D = 1$ and $D = 2$ improves the DSC performance for all four brain structures and statistically significant changes can be observed for the germinal matrix and white matter. Moreover, the standard error for GMAT is reduced almost by half, from 0.131 for $D = 0$ to 0.069 for $D = 1$, thanks to more consistent performance of GMAT segmentation across all gestational ages.

4 Conclusions

We presented an approach to building a spatio-temporal model of tissue distributions in the developing fetal brain which can incorporate both developed tissues as well as transient tissue classes such as the germinal matrix. The model is created from a set of manual segmentations through groupwise registration and voxelwise non-linear modeling of tissue class membership, that allows us to represent the appearance as well as disappearance of brain structures over time.

Based on the spatio-temporal model, we generated a set of age-specific tissue probability maps and used them as sources of tissue priors for atlas-based EM

segmentation. Experiments performed using clinical MR scans of 10 fetal subjects demonstrated that the use of age-specific atlases improves the performance of automatic brain segmentation as measured by the average DSC values.

The presented study will be extended by spatio-temporal modeling of tissue contrast changes in fetal MR images. This will allow us to create age-specific synthetic MR images that can be used as targets for registration of new subjects to the average shape space before they can be segmented with age-specific probabilistic atlases generated from the spatio-temporal model.

Acknowledgments. This work was primarily funded by NIH Grant R01 NS 55064. The work of Orit A. Glenn and imaging studies were funded by NIH Grant K23 NS 52506.

References

1. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic segmentation of MR images of the developing newborn brain. Med. Image Anal. 9(5), 457–466 (2005)
2. Xue, H., Srinivasan, L., Jiang, S., Rutherford, M., Edwards, A.D., Rueckert, D., Hajnal, J.V.: Automatic segmentation and reconstruction of the cortex from neonatal MRI. Neuroimage 38(3), 461–477 (2007)
3. Murgasova, M., Dyet, L., Edwards, D., Rutherford, M., Hajnal, J., Rueckert, D.: Segmentation of brain MRI in young children. Acad. Radiol. 14(11), 1350–1366 (2007)
4. Yoon, U., Fonov, V.S., Perusse, D., Evans, A.C.: The effect of template choice on morphometric analysis of pediatric brain data. Neuroimage 45(3), 769–777 (2009)
5. Studholme, C., Cardenas, V.A., Weiner, M.W.: Multiscale image and multiscale deformation of brain anatomy for building average brain atlases. In: Medical Imaging: Image Processing. In: Proc. SPIE, vol. 4322 (2001)
6. Prayer, D., Kasprian, G., Krampl, E., Ulm, B., Witzani, L., Prayer, L., Brugger, P.C.: MRI of normal fetal brain development. Eur. J. Radiol. 57(2), 199–216 (2006)
7. Kim, K., Hansen, M.F., Habas, P.A., Rousseau, F., Glenn, O.A., Barkovich, A.J., Studholme, C.: Intersection-based registration of slice stacks to form 3D images of the human fetal brain. In: Proc. 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 1167–1170 (2008)
8. Studholme, C., Hill, D.L.G., Hawkes, D.J.: An overlap invariant entropy measure of 3D medical image alignment. Pattern Recognit. 32(1), 71–86 (1999)
9. Pohl, K.M., Fisher, J., Shenton, M., McCarley, R.W., Grimson, W.E.L., Kikinis, R., Wells, W.M.: Logarithm odds maps for shape representation. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 955–963. Springer, Heidelberg (2006)
10. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based bias field correction of MR images of the brain. IEEE Trans. Med. Imaging 18(10), 885–896 (1999)
11. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Trans. Med. Imaging 18(10), 897–908 (1999)
12. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology 26(3), 297–302 (1945)

Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets

Stanley Durrleman[1,2], Xavier Pennec[1], Alain Trouvé[2], Guido Gerig[3], and Nicholas Ayache[1]

[1] INRIA - Asclepios Team-Project, Sophia Antipolis, France
[2] Centre de Mathématiques et Leurs Applications (CMLA), ENS-Cachan, France
[3] Scientific Computing and Imaging (SCI) Institute, University of Utah, USA

Abstract. We propose a new methodology to analyze the anatomical variability of a set of longitudinal data (population scanned at several ages). This method accounts not only for the usual 3D anatomical variability (geometry of structures), but also for possible changes in the dynamics of evolution of the structures. It does not require that subjects are scanned the same number of times or at the same ages. First a regression model infers a continuous evolution of shapes from a set of observations of the same subject. Second, spatiotemporal registrations deform jointly (1) the geometry of the evolving structure via 3D deformations and (2) the dynamics of evolution via time change functions. Third, we infer from a population a prototype scenario of evolution and its 4D variability. Our method is used to analyze the morphological evolution of 2D profiles of hominids skulls and to analyze brain growth from amygdala of autistics, developmental delay and control children.

1 Methodology for Statistics on Longitudinal Data

Many frameworks has been already proposed in medical imaging to analyze the anatomical variability of 3D structures like images, curves or surfaces. Less attention has been paid to the variability of longitudinal data (several subjects scanned several times). In [1], the evolution between two shapes is modeled by a geodesic deformation, which cannot be used for more than two data per subjects. In [2], shape growth is measured via the evolution of extracted features like volumes, shape or pose parameters. In [3,4], a temporal regression is proposed globally for a population, but this does not allow inter-subject comparisons. In cardiac motion analysis [5,6], spatiotemporal registration relies on 3D-registrations between images of the same moment of the cardiac cycle and between two consecutive time-points. These works rely on time-point correspondence and do not call the labels of the time-points into question. By contrast, in longitudinal studies, subjects are scanned at ages which do not necessarily correspond. Moreover, evolutions may be delayed or advanced within a population, a key feature that we precisely aim at detecting. In [7,8], deformation

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 297–304, 2009.

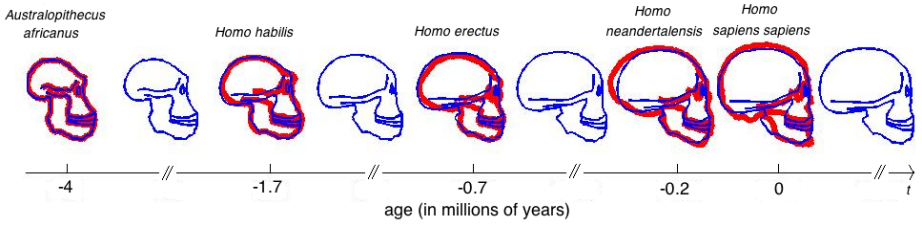


Fig. 1. Skull profile of five hominids (in red). The regression model estimates a continuous evolution (in blue) of the Australopithecus, which closely matches the data.

of cardiac motion are proposed both in space and time but they require a fine temporal sampling of the motion, whereas only few acquisitions per subjects are available in most longitudinal studies. In this paper, we propose to use a regression model to estimate a continuous evolution from data sparsely distributed in time and spatiotemporal deformations which register jointly both the 3D geometry and the scenario of evolution. Geometrical data are modeled as currents to avoid assuming point correspondence between structures. Large deformations are used which gives a rigorous framework for statistics on deformations and atlas construction [9,10,11]. From longitudinal data, we estimate consistently the most likely scenario of evolution and its spatiotemporal variability within the population.

In this paper, we call longitudinal data a set of geometrical data (curves or surfaces, called here shapes), acquired from different subjects scanned at several time-points. We assume that the successive data of a given subject are temporal samples of a continuous evolution. We propose therefore a **regression model** which computes a continuous evolution which matches the data of the subject at the corresponding time points (Fig. 1). This continuous evolution allows us to compare two subjects at a given age, even if one subject has not been scanned at this age. We can also analyze how the shape varies near this age to detect possible developmental delays. We define then the **spatiotemporal deformation** of a continuous evolution, which consists of two deformations: (1) a morphological deformation (of the 3D space) which changes the geometry of every frame of the evolution *independently of the time point* and (2) a time change function (deformation of the time interval) which changes the dynamics of the evolution *without changing the geometry of shapes*. To avoid time-reversal, the time change function must be smooth and order preserving: it is a diffeomorphism of the time interval of interest. A **4D registration** between two subjects looks for the most regular spatiotemporal deformation, such that the deformation of the continuous evolution inferred from the first subject maps the successive target data (Fig. 2). Eventually, we use this 4D registration framework to estimate a **spatiotemporal atlas** from a population, based on an 4D extension of the statistical model of [11]. We look for a template and a continuous evolution of this template (called mean scenario of evolution), so that data of each subject

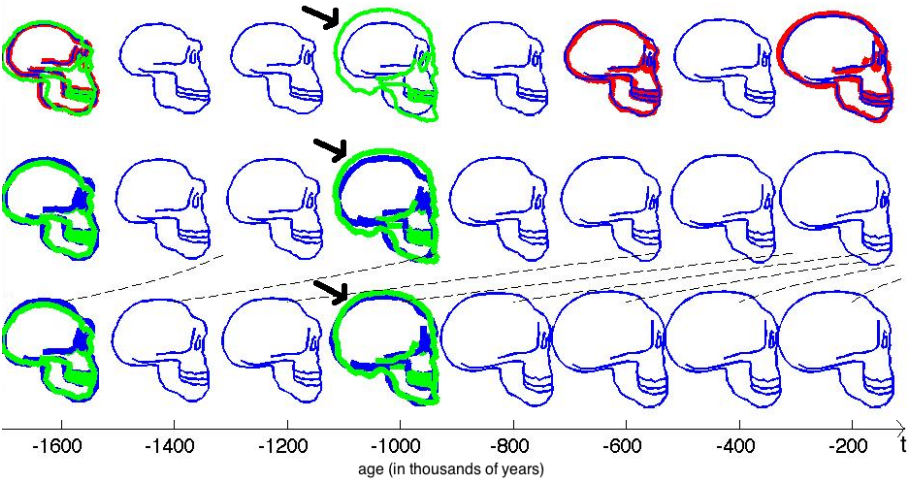


Fig. 2. Registration of the evolution {Homo habilis-erectus-neandertalensis} (in red) to the evolution {Homo erectus-sapiens sapiens} (in green), shifted to start at the same time. **Top row:** Regression of the source data (red) gives the continuous evolution in blue. **Middle row:** The geometrical part ϕ is applied to each blue frame. This shows morphological changes: the skull is larger, rounder and the jaw less prominent. **Bottom row:** The time change function ψ is applied to the evolution of the second row. The blue shapes are moved along the time axis (as shown by dashed black lines), but they are not deformed. Black arrows show that a better alignment is achieved when one accounts both for morphological changes *and* a change of the evolution speed.

are temporal samples of a spatiotemporal deformation of the mean scenario. A Maximum A Posteriori estimation enables to estimate consistently the template, the mean scenario and the spatiotemporal deformations of this mean scenario to each subject.

We present the regression model in Sec. 2, 4D registration in Sec. 3 and atlas construction in Sec. 4. We focus here on the methodology and refer the reader to [12] for more computational details. In Sec. 5, we apply our method on 2D profiles of hominids skulls and amygdala of children scanned at 2 and 4 years.

2 Regression Model for Shape Evolution

We want to fit a continuous shape evolution to a set of shapes (S_i) of the same subject acquired at different time points (t_i). Without loss of generality, we can assume that $t_{\min} = 0$ and $t_{\max} = T$. This evolving shape is equal to the baseline M_0 at time $t = 0$, which may be the earliest shape of this subject or a template as in Sec. 4. The evolution has the form: $M_t = \chi_t(M_0)$ where t varies continuously in the time interval $[0, T]$. For each t, χ_t is a diffeomorphism of the 3D space, such that $\chi_0 = \mathrm{Id}$ (which leads to $\chi_0(M_0) = M_0$). The regression (M_t) must match

the observation S_i at the time-points t_i, while a rigidity constraint controls the regularity of the regression. This is achieved by minimizing:

$$J(\chi) = \sum_{t_i} \mathrm{d}(\chi_{t_i}(M_0), S_i)^2 + \gamma_\chi \mathrm{Reg}(\chi) \tag{1}$$

where d is a similarity measure between shapes, $\mathrm{Reg}(\chi)$ a regularity term and γ_χ a trade-off between regularity and fidelity to data. Among other possible choices, we use here the large deformations of [13], and model curves or surfaces as currents [14,10]. Therefore, d is the distance between currents and χ is the solution of the partial differential equation : $\partial\chi_t(x)/\partial t = v_t^\chi(\chi_t(x))$ with initial condition $\chi_0 = \mathrm{Id}$. The regularity term $\mathrm{Reg}(\chi)$ is given by $\int_0^T \|v_t^\chi\|^2 \, dt$. In the setting of [14], J depends on time-dependent momenta $\alpha_k(t)$ at each samples of M_t's, which are used as variables for the gradient descent [13,12].

As a result, for all t, χ_t depends on all the constraints in the past and future. This differs from pairwise registrations between consecutive time-points. The function χ is piecewise geodesic and can be extended at all times by assuming $v_t^\chi = 0$ (and hence χ_t constant) outside $[0, T]$. This is useful to compare this evolution with another subject which may have data outside the time interval.

3 Spatiotemporal Pairwise Registration

We assume now that we have successive shapes for the source subject $(S(t_i))_i$ and for the target $(T(t_j))_j$. As in Sec 2, we perform a regression on the source shapes which leads to a continuous evolution $S(t)_{t\in[0,T]}$. Our goal is to find a diffeomorphism of the 3D space ϕ and a diffeomorphism of the time-interval ψ which deform the source evolution $S(t)$ into $S'(t) = \phi(S(\psi(t)))$ such that $S'(t_j)$ match $T(t_j)$. Thanks to the regression function, no correspondence is needed between the time points t_i and t_j. Formally, we minimize:

$$J(\phi, \psi) = \sum_{t_j} \mathrm{d}(\phi(S(\psi(t_j))), T_{t_j})^2 + \gamma_\phi \mathrm{Reg}(\phi) + \gamma_\psi \mathrm{Reg}(\psi) \tag{2}$$

The spatial (resp. temporal) deformation ϕ (resp. ψ) is solution at parameter $u = 1$ of the flow equation $\partial_u \phi_u(x) = v_u^\phi(\phi_u(x))$ (resp. $\partial_u \psi_u(t) = v_u^\psi(\psi(t))$). The norm of the speed vector fields v_u^ϕ and v_u^ψ integrated for $u \in [0, 1]$ defines the regularity terms $\mathrm{Reg}(\phi)$ and $\mathrm{Reg}(\psi)$ respectively. As in Sec 2, the geometrical (resp. temporal) deformation is parametrized by momenta α (resp. β) at the points of $S(t_j)$ (resp. at the t_j), which are used as variables for the gradient descent. $\nabla_\alpha J(\alpha, \beta)$ is computed as for the registration of the collection of $\phi(S(\psi(t_j)))$ to the (T_{t_j}) [14,10]. To compute $\nabla_\beta J$, we need to derive the matching term with respect to the time-points $t'_j = \psi(t_j)$ [12]. We use here a centered Euler scheme: $\nabla_t \|\phi(S(t)) - T\|^2 \sim \langle \phi(S(t)) - T, \phi(S(t + \delta t)) - \phi(S(t - \delta t)) \rangle / \delta t$, where δt is the discrete time step. We used here the fact that the shapes are embedded with a vector space (the space of currents) provided with an inner product. Note that we minimize J with respect to the geometrical and the temporal parameters *jointly*. We do not performed alternated minimization.

4 Spatiotemporal Atlas Construction

We assume now that we have a set of N subjects (S^i), provided each with temporal observations $(S^i(t^i_j))_j$. We are looking for a template M_0 and a mean scenario of evolution of this template $M(t) = \chi_t(M_0)$, such that the observations correspond to particular moments of a spatiotemporal deformation of the mean scenario. This means that $\phi^i(M(\psi^i(t^i_j)))$ match $S^i(t^i_j)$ for each subject i and time t^i_j. Maximum A Posteriori estimation in the same setting as in [11], leads to the minimization of $J(\psi^i, \phi^i, \chi, M_0) =$

$$\sum_{i=1}^{N} \left\{ \sum_{t^i_j} \mathrm{d}(\phi^i(\chi_{\psi^i(t^i_j)} M_0), S^i(t^i_j))^2 + \gamma_\phi \mathrm{Reg}(\phi^i) + \gamma_\psi \mathrm{Reg}(\psi^i) + \gamma_\chi \mathrm{Reg}(\chi) \right\}$$

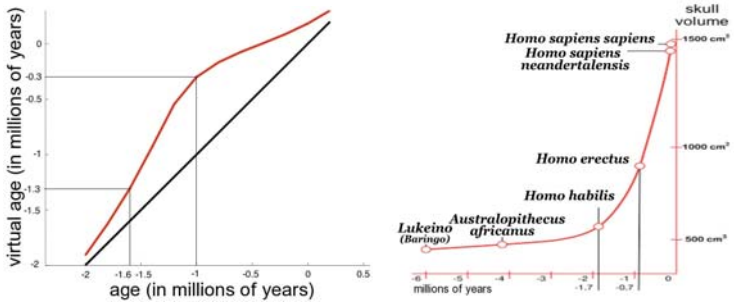
We perform a 3 step alternated minimization. If the template M_0 and the regression χ are fixed, the minimum is achieved for N registrations of the mean scenario $\chi_t(M_0)$ to each subject's set of data $S^i(t^i_j)$, as in Sec. 3. If we fix the N spatiotemporal deformations (ϕ^i, ψ^i) and the regression χ, we need to minimize $\sum_{i,j} \mathrm{d}(\Phi_{i,j}(M_0), S^i(t^i_j))^2$, where the $\Phi_{i,j} = \phi^i \circ \chi_{\psi^i(t^i_j)}$ are 3D-diffeomorphisms. This is exactly the estimation of an unbiased template in the setting of [11], when the deformations are given by $\Phi_{i,j}$. When the template M_0 and the N spatiotemporal deformations (ϕ^i, ψ^i) are fixed, we need to minimize $\sum_{i,j} \mathrm{d}(\phi^i(\chi_{\psi^i(t^i_j)} M_0), S^i(t^i_j))^2 + \gamma_\chi \mathrm{Reg}(\chi)$. This is not the regression problem stated in Sec. 2 because of the deformation ϕ^i in the matching term. To turn it into regression, we approximate the matching term $\mathrm{d}(\phi^i(\chi_{\psi^i(t^i_j)} M_0), S^i(t^i_j))$ by $\mathrm{d}(\chi_{\psi^i(t^i_j)}(M_0), (\phi^i)^{-1}(S^i(t^i_j)))$ (subject's shapes are matched back to the mean anatomy). This approximation is valid only for diffeomorphisms ϕ^i whose Jacobian is close to the identity. To initialize the minimization, we set M_0 as the mean current of the earliest data and set the diffeomorphisms χ, ϕ^i, ψ^i to identity.

5 Numerical Experiments

Experiments on 2D curves relates to Sec 2 and 3, those on surfaces to Sec. 4.

Evolution of 2D Curves. We have five 2D-profiles of hominids skulls which consist of six lines each (source: www.bordalierinstitute.com). Our regression framework infers a continuous evolution from the Australopithecus to the Homo sapiens sapiens which matches the intermediate stages of evolution in Fig. 1.

Then, we register the evolution {Homo habilis-Homo erectus-Homo neandertalensis} to the evolution {Homo erectus-Homo sapiens sapiens} in Fig 2. The geometrical deformation shows that during the later evolution the jaw was less prominent and the skull larger and rounder than during the earlier evolution. The time change function shows that the later evolution occurs at a speed 1.66 times faster than the earlier evolution. This value is compatible with the growth speed of the skull during these periods (See Fig. 3): between Homo erectus and sapiens the skull volume growths at $(1500 - 900)/0.7 = 860\mathrm{cm}^3/10^6\mathrm{years}$, whereas



a- time change function b- skull volume evolution

Fig. 3. A- time change function $\psi(t)$ of the registration in Fig. 2 (in black the reference $\psi(t) = t$). The slope of the curve measures an acceleration between evolutions, which is compatible with the growth of skull volume in b (source: www.bordalierinstitute.com).

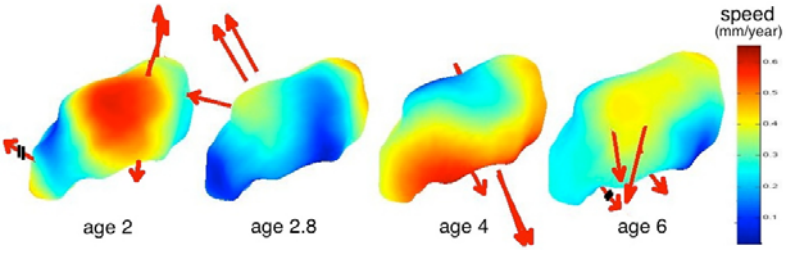


Fig. 4. Mean Scenario of the right Amygdala (right lateral part). Arrows measures the differences between age t+0.2 and age t in the space of currents as in [16]. From age 2 to 2.8, the evolution is mainly a torque at the posterior part; then the structure becomes thicker, mostly at the superior part between age 2.8 and 4 and at the inferior between age 4 and 6; from age 6 the evolution is a mainly a torque at the anterior part.

between Homo habilis and neandertalensis, it growths at $(1500 - 600)/1.7 = 530\text{cm}^3/10^6\text{years}$, namely 1.62 times faster.

Evolution of 3D Surfaces. We use here meshes of amygdala of the right hemisphere from 4 autistics, 4 developmental delay and 4 control children scanned twice [15]. Age distribution is shown in Fig 5-a. From these data registered rigidly, we infer a template, a mean scenario of evolution of this template and the spatiotemporal evolution of this mean scenario to each subject. In the setting of [13,10], the diffeomorphisms are controlled by the standard deviation of Gaussian kernel set to 15 mm for χ, ϕ and 1 year for ψ; the typical scale on currents is set to 3 mm. Trade-off γ_χ, γ_ϕ were set to 10^{-3} and γ_ψ to 10^{-6}. An amygdala is typically 10 mm large. The discrete time step is set to 0.2 years.

By inspection of the companion movie of the mean scenario, one distinguishes 4 phases during growth (See also Fig. 4). Preliminary tests do not show correlations between the morphological deformations and the pathology. From the time change functions shown in Fig. 5, we cannot conclude that a subject with pathology is systematically delayed or advanced compared to controls, even at a given age. However, the curves show that the growth speed seems to follow the

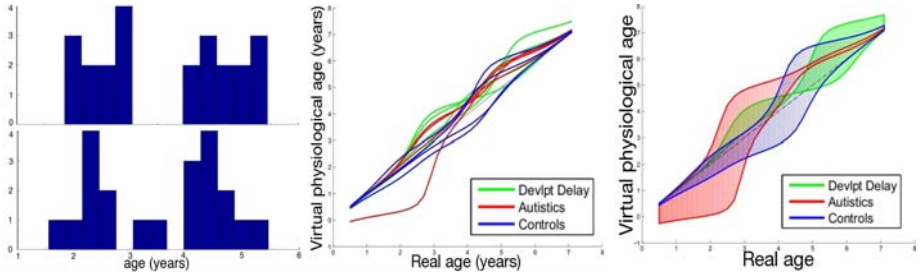


Fig. 5. Temporal deformation of the mean scenario Left: distribution of original (top) and registered (bottom) ages. Middle: time change functions for the 12 subjects. Right: First mode of variation at $\pm\sigma$ of the time change functions for each class. Autistics and controls show the same evolution pattern, but shifted in time.

same pattern, mainly an acceleration between age 2.5 and 3.5 for the autistics and between age 4 and age 5 for controls. The developmental delay also have such pattern but it occurs at a very variable age. These results suggest that the discriminative information between classes might not be inferred from the anatomical variability at a given age, but rather from variations of the growth process. These results, however, must be strengthen using larger database. The more time-points per subjects, the more constrained the mean scenario estimation. The more subjects, the more robust the statistics.

6 Discussion and Conclusion

In this paper, we present a generic framework to analyze variability of longitudinal data. A regression model fits a continuous evolution to successive data of one subject. 4D registrations decompose the difference between two sets of longitudinal data into a geometrical deformation and a change of the dynamics of evolution. The more acquisitions per subjects, the more constrained this decomposition. However, no constraint is imposed in terms of number and correspondence of measurement points across subjects. These pairwise registrations are used for group-wise statistics: ones estimates consistently a template, the mean evolution of this template and the spatiotemporal variability of this evolution in the population. Then, statistical measures can be derived, like the first mode of temporal deformation in Fig. 5. Further experiments have still to be performed to give more quantitative measures of variability. However, these first results suggest that pathologies might be characterized more by a particular scenario of evolution than by the anatomy at a given age. Our methodology can be used therefore to drive the search of new anatomical knowledge and to give characterization of pathologies in terms of organ growth scenario. This may be applied to the study of degenerative diseases or cardiac motion disorders.

Acknowledgments. This work was partially supported the European IP project Health-e-child (IST-2004-027749) and Microsoft Research.

References

1. Qiu, A., Younes, L., Miller, M., Csernansky, J.: Parallel transport in diffeomorphisms distinguishes the time-dependent pattern of hippocampal surface deformation due to healthy aging and the dementia of the alzheimer's type. NeuroImage 40, 68–76 (2008)
2. Gorczowski, K., Styner, M., Jeong, J.Y., Marron, J.S., Piven, J., Hazlett, H.C., Pizer, S.M., Gerig, G.: Statistical shape analysis of multi-object complexes. Transactions on Pattern Analysis and Machine Intelligence (to appear, 2009)
3. Khan, A., Beg, M.: Representation of time-varying shapes in the large deformation diffeomorphic framework. In: Proc. of ISBI 2008, pp. 1521–1524 (2008)
4. Davis, B., Fletcher, P., Bullitt, E., Joshi, S.: Population shape regression from random design data. In: Proc. of ICCV 2007, pp. 1–7 (2007)
5. Chandrashekara, R., Rao, A., Sanchez-Ortiz, G.I., Mohiaddin, R.H., Rueckert, D.: Construction of a statistical model for cardiac motion analysis using nonrigid image registration. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 599–610. Springer, Heidelberg (2003)
6. Peyrat, J.M., Delingette, H., Sermesant, M., Pennec, X., Xu, C., Ayache, N.: Registration of 4D Time-Series of Cardiac Images with Multichannel Diffeomorphic Demons. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 972–979. Springer, Heidelberg (2008)
7. Declerck, J., Feldman, J., Ayache, N.: Definition of a 4D continuous planispheric transformation for the tracking and the analysis of LV motion. Medical Image Analysis 4(1), 1–17 (1998)
8. Perperidis, D., Mohiaddin, R.H., Rueckert, D.: Spatio-temporal free-form registration of cardiac MRI sequences. Medical Image Analysis 9(5), 441–456 (2005)
9. Vaillant, M., Miller, M., Younes, L., Trouvé, A.: Statistics on diffeomorphisms via tangent space representations. NeuroImage 23, 161–169 (2004)
10. Durrleman, S., Pennec, X., Trouvé, A., Thompson, P., Ayache, N.: Inferring brain variability from diffeomorphic deformations of currents: an integrative approach. Medical Image Analysis 12(5), 626–637 (2008)
11. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Statistical models of sets of curves and surfaces based on currents. Medical Image Analysis (to appear, 2009)
12. Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., Ayache, N.: Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. Research Report 6952, INRIA (June 2009)
13. Miller, M.I., Trouvé, A., Younes, L.: On the metrics and Euler-Lagrange equations of computational anatomy. Annual Review of Biomed. Eng. 4, 375–405 (2002)
14. Vaillant, M., Glaunès, J.: Surface matching via currents. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 381–392. Springer, Heidelberg (2005)
15. Hazlett, H., Poe, M., Gerig, G., Smith, R., Provenzale, J., Ross, A., Gilmore, J., Piven, J.: Magnetic resonance imaging and head circumference study of brain size in autism. The Archives of General Psychiatry 62, 1366–1376 (2005)
16. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Sparse approximation of currents for statistics on curves and surfaces. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 390–398. Springer, Heidelberg (2008)

On the Manifold Structure of the Space of Brain Images

Samuel Gerber, Tolga Tasdizen, Sarang Joshi, and Ross Whitaker

Scientific Computing and Imaging Institute, University of Utah, USA
sgerber@cs.utah.edu

Abstract. This paper investigates an approach to model the space of brain images through a low-dimensional manifold. A data driven method to learn a manifold from a collections of brain images is proposed. We hypothesize that the space spanned by a set of brain images can be captured, to some approximation, by a low-dimensional manifold, i.e. a parametrization of the set of images. The approach builds on recent advances in manifold learning that allow to uncover nonlinear trends in data. We combine this manifold learning with distance measures between images that capture shape, in order to learn the underlying structure of a database of brain images. The proposed method is generative. New images can be created from the manifold parametrization and existing images can be projected onto the manifold. By measuring projection distance of a held out set of brain images we evaluate the fit of the proposed manifold model to the data and we can compute statistical properties of the data using this manifold structure. We demonstrate this technology on a database of 436 MR brain images.

1 Introduction

Recent research in the analysis of populations of brain images shows a progression: from single templates or atlases [1], to multiple templates or stratified atlases [2], mixture models [3] and template free methods [4,5,6] that rely on a sense of *locality* in the space of all brains. This progression indicates that the space of brain MR images has a structure that might also be modeled by a relatively low-dimensional manifold as illustrated by Figure 1. The aim of this paper is to develop and demonstrate the technology to learn the manifold structure of sets of brain MR images and to evaluate how effective the learned manifold is at capturing the variability of brains.

Manifold learning [7] refers to the task of uncovering manifolds that describe scattered data. In some applications this manifold is considered a generative model, analogous to a Gaussian mixture model. In this context, we assume that the data is sampled from a low-dimensional manifold embedded in a high-dimensional space, with the possibility of noise that sets data off the surface. For this work, we consider the space of all images which can be represented as smooth functions. Virtually all manifold learning techniques published to date assume that the the low-dimensional manifold is embedded in a Euclidean space.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 305–312, 2009.

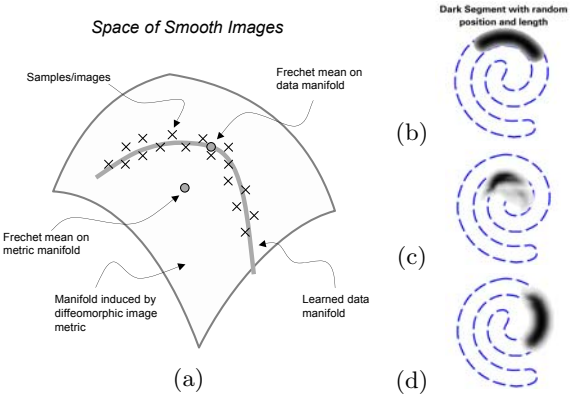


Fig. 1. (a) Illustration of image data on a low-dimensional manifold embedded in a diffeomorphic space. (b) A set of images consists of random length/position segments form a spiral. (c) The Fréchet mean in the diffeomorphic space is not like any example from the set. (d) Fréchet mean on data-driven manifold reflects the image with average parameter values.

Nearby samples lie in the tangent space of the manifold, and thus their differences can be evaluated by Euclidean distance in the ambient space. The space of brain images on the other hand does not fit directly into this paradigm. A great deal of research on brain image analysis shows that the L_2 distance is not suitable for measuring shape changes in images [8], but that the metric for comparing brain images should account for deformations or shape differences between images. For example, computational anatomy, used for population analysis and atlas building, is based on a metric between images derived from coordinate transformations [2,9,3].

The low-dimensional manifold of brain images we aim to learn is embedded not in Euclidean space, but in the space of images with a metric based on coordinate transformations. For this work we adapt the image metric based on diffeomorphic coordinate transformations [10,11,12] to manifold learning. Often the stratification induced by the diffeomorphic image metric is described as a *manifold*—in this paper we refer to the manifold of brain images as described by the data. Our hypotheses are that the space of brain images is some very small subspace of images that are related by diffeomorphisms, that this subspace is not linear, and that we can learn some approximation of this space through a generalization of manifold learning that accounts for these diffeomorphic relationships. Figure 1 illustrates these concepts on a simple example.

A manifold learning algorithm of particular interest to this work is isomap [13]. Isomap is based on the idea of approximating geodesic distances by the construction of piecewise linear paths between samples. The paths are built by connecting nearest neighbors, and the geodesic distance between two points is approximated on the the linear segments between nearest neighbors. Thus, isomap requires only distances between nearby data points to uncover manifold structure in data sets.

The reliance on only nearest neighbor distances is important for this paper. The tangent space to the space of diffeomorphic maps is the set of smooth vector fields. Thus, if the samples from the manifold are sufficiently dense, we can compute the distances in this tangent space, and we need only to compute elastic deformations between images.

Isomap, and several other manifold learning algorithms, assign parameters to data points that represent coordinates on the underlying manifold. This paper introduces several extensions to this formulation, which are important for the analysis of brain images. First is an explicit representation of the manifold in the ambient space (the space of smooth functions). Thus, given coordinates on the manifold, we can construct brain images that correspond to those coordinates. We also introduce a mechanism for mapping previously unseen data into the manifold coordinate system. These two explicit mappings allow to project images onto the manifold. Thus we can measure the distance from each image to the manifold (projected image) and quantitatively evaluate the efficacy of the learned manifold. In comparison with previous work, on brain atlases for example, this work constructs, from the data itself, a parametrized hyper-surface of brain images, which represents a *local atlas* for images that are nearby on the manifold.

2 Related Work

The tools for analyzing or describing sets of brain image demonstrate progressively more sophisticated models. For instance, unbiased atlases are one mechanism for describing a population of brains [14,15,16]. Blezek et al. [2] propose a stratified atlas, in which they use the mean shift algorithm to obtain multiple templates and shows visualizations that confirm the clusters in the data. In [3] the OASIS brain database is modeled through a mixture of Gaussians. The means of the Gaussians are a set of templates used to describe the population. Instead of assuming that the space of brain images forms clusters, we postulate that the space of brains can be captured by a continuous manifold.

An important aspect of our work is the ability to measure image differences in a way that captures shape. It is known that the L_2 metric does not adequately capture shape differences [8]. There are a variety of alternatives, most of which consider coordinate transformations instead of, or in addition to, intensity differences. A large body of work [10,11,12] has examined distances between images based on high-dimensional image warps that are constrained to be diffeomorphisms. This metric defines a infinite dimensional manifold consisting of all shapes that are equivalent under a diffeomorphism. Our hypothesis, however, is that the space of brains is essentially of significantly lower dimension.

Several authors [17,9,6] have proposed kernel-based regression of brain images with respect to an underlying parameter, such as age. The main distinction of the work in this paper is that the underlying parametrization is *learned* from the image data. Our interest is to uncover interesting structures from the image data and sets of parameters that could be compared against underlying clinical variables.

Zhang et al. use manifold learning, via isomap, for medical image analysis, specifically to improve segmentation in cardiac MR images [18]. Rohde et al. [19] use isomap in conjunction with large deformation diffeomorphisms to embedded binary images of cell nuclei in a similar fashion to the proposed approach. In addition to the embedding we provide a generative model that allows to quantitatively evaluate the manifold fit.

3 Formulation

We begin with a description of the image metric between nearest neighbors in the space of smooth images. A diffeomorphic coordinate transformation between two images is $\phi(x,1)$, where $\phi(x,t) = x + \int_0^t v(\phi(x,\tau),\tau)d\tau$, and $v(x,t)$ is a smooth, time varying vector field. The diffeomorphic framework includes a metric on the diffeomorphic transformation $||\phi(x,t)||_L = \int_0^t ||v(x,\tau)||_L d\tau$ which induces a metric d between images y_i and y_j:

$$\begin{aligned} d(y_i, y_j) = \min_v & \int_0^1 ||v(x,\tau)||_L d\tau \\ \text{subject to } & \int_\Omega ||y_i(x+\phi(x,1)) - y_j(x))||_2^2 dx = 0 \end{aligned} \tag{1}$$

The metric prioritizes the mappings and, with an appropriate choice of the differential operator L in the metric, ensures smoothness. We introduce the constraint that the transformation must provide a *match* between the two images:

$$||y_i(\phi(x,1)) - y_j(x)|| = \left(\int_\Omega (y_i(\phi(x,1)) - y_j(x))^2 dx \right)^{\frac{1}{2}} < \epsilon. \tag{2}$$

where ϵ allows for noise in the images.

For two images that are very similar, ϕ and v are small, and because the velocities of the geodesics are smooth in time [20], we can approximate the integrals for the coordinate transform and geodesic distance:

$$\begin{aligned} \phi(x,1) \approx v(x,0) = v(x), \text{ and } & \mathrm{d(y_i, y_j)} \approx \min_{\mathrm{v}} ||\mathrm{v(x)}||_{\mathrm{L}}, \\ & \text{subject to } \int_\Omega ||y_i(x+v(x)) - y_j(x))||_2^2 dx < \epsilon \end{aligned}. \tag{3}$$

Thus, for small differences in images the diffeomorphic metric is approximated by a smooth displacement field. In this paper we use the operator $L = \alpha \mathbf{I} + \nabla$, where α is a free parameter and the resulting metric is $||v(x)||_L = \|Lv(x)\|_2$. To minimize deformation metric for a pair of discrete images, we use a gradient descent. The first variation of (3) results in a partial differential equation, which we solve with finite forward differences to an approximate steady state. For the constraint, we introduce a penalty on image residual with an additional parameter λ, which we tune in steady state until the residual condition in (2) is satisfied or until the deformation metric exceeds some threshold that disqualifies that pair of images as nearest neighbors. We use a multiresolution, coarse to fine, optimization strategy to avoid local minima.

Next we present a formulation for representing the structure of the manifold in the ambient space and for mapping unseen data onto this intrinsic coordinate

system. First, we propose the construction of an explicit mapping $f : \mathcal{P} \to \mathcal{A}$ from the space of manifold parameters $\mathcal{P}$ to the high dimensional ambient space $\mathcal{A}$. Let $X = \{x_1, \ldots, x_n\}$ be the parameter values assigned to the image data sets $Y = \{y_1, \ldots, y_n\}$; isomap gives the discrete mapping $x_i = \rho(y_i)$. Inevitably there will be a distribution of brain images away from the manifold, and the manifold should be the expectation [21] of these points in order to alleviate noise and capture the overall trend in the data. That is $f(x) = E(Y|\rho(Y) = x)$. In the discrete setting the conditional expectation can be approximated with Nadaraya-Watson kernel regression:

$$f(x) = \arg\min_y \frac{\sum_{i \in X_{nn}(x)} K(x, x_i)\tilde{d}(y, y_i)}{\sum_{i \in X_{nn}(x)} K(x, x_i)}, \tag{4}$$

which we compute, in the context of diffeomorphic image metrics using the method of [9], which iteratively updates $f(x)$ and the deformation to $f(x)$ from the nearest neighbors starting with identity transformations. This kernel regression requires only the nearest neighbors $X_{nn}(x)$ of $x_i \in X$. This constrains the regressions to images similar in shape since locality in X implies locality in Y. Using this formulation, we can compute an image for any set of manifold coordinates, and thus we have an explicit parametrization of the manifold.

For the assignment of manifold parameters to new, unseen images we use the same strategy. We represent this mapping as a continuous function on the ambient space, and we compute it via a regression on parameters given by isomap

$$\rho'(y) = \frac{\sum_{i \in Y_{nn}(y)} K(y, y_i)x_i}{\sum_{i \in Y_{nn}(y)} K(y, y_i)}, \tag{5}$$

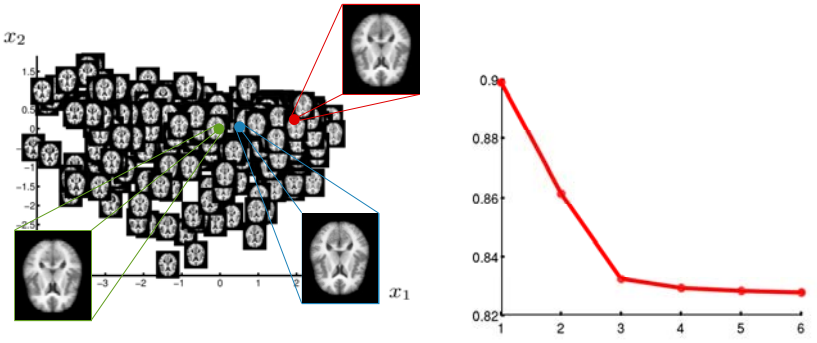
with $Y_{nn}(y)$ the nearest neighbors of y. The projection of a new image onto the manifold is the composition of these mappings $p(y) = f(\rho'(y))$.

For K we use a Gaussian kernel for the mappings with a bandwidth selected based on average nearest neighbor distances. The number of nearest neighbors for the regression is selected based on the resulting bandwidth for the kernel K, such that all points within three standard deviations are included.

4 Results

In section 1 we illustrated the idea of the paper on a simple examples on 2D images of spiral segments. The image data set used consists of 100 images of segments with varying length and location of the spiral in Figure 1. Figure 2 shows images constructed by the proposed approach by sampling the learned manifold representation of the image data. Thus the images depict samples on the manifold embedded in the ambient space. Figure 1 also shows the Fréchet means for the diffeomorphic space and for the manifold learned from the data.

Fig. 2. Reconstructed images along the first dimension of the manifold learned from spiral segments as illustrated in Figure 1



(a) (b)

Fig. 3. (a)2D parametrization of OASIS brain MRI. The insets show the mean (green), median (blue) and mode (red) of the learned manifold and the corresponding reconstructed images. (b) Reconstruction errors against manifold dimensionality.

We apply the proposed approach to the open access series of imaging studies (OASIS) cross-sectional MRI data set. The images are gain-field corrected and atlas registered. We use 380 of the 436 images to learn the manifold and evaluate reconstruction errors on the left out 56 images.

Figure 3 (a) shows axial slice 80 for a 2D parametrization (manifold coordinates) obtained by the proposed method. A visual inspection reveals that the learned manifold detects the change in ventricle size as the most dominant parameter (horizontal axis). It is unclear if the second dimension (vertical axis) captures a global trend. Figure 3 (b) shows reconstruction errors on the held out images against the dimensionality of the learned manifold. The reconstruction error is measured as the mean of the distances between the original brain images and their projection on to the learned manifold scaled by the average nearest neighbor distance, i.e. error $= \frac{\sum_i d(f(\rho'(y_i)),y_i)}{\sum_i d(nn(y_i),y_i)}$. The reconstruction errors are smaller than the average one nearest neighbor distance. An indication that the learned manifold accurately captures the data. The reconstruction errors suggest that the data set can be captured by a 3D manifold. We do not postulate that the space of brains is captured by a 3D manifold. The approach learns a manifold from the available data and thus it is likely that given more samples we can learn a higher dimensional manifold for the space of brains.

Figure 4 shows axial slices of brain images generated with the proposed method on a regularly sampled grid on the 2D representation shown in figure 3(c), i.e. we have sampling of the learned brain manifold. The first

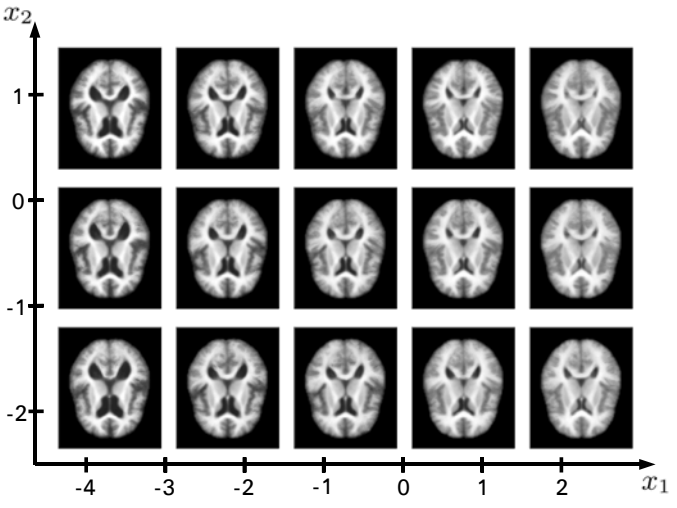


Fig. 4. Reconstructions on a grid on the 2D representation shown in figure 3(c)

dimension (x_1) clearly shows the change in ventricle size. The second dimension (x_2) is less obvious. A slight general trend observable from the axial slices seems to be less gray and white matter as well as a change in lateral ventricle shape (from elongated to more circular).

The method is computationally expensive because of the pairwise distance computations, each requiring an elastic image registration. The registration takes with our multiresolution implementation about 1 minute on a $128 \times 128 \times 80$ volume. Pairwise distances computations for the OASIS database running on a cluster of 50, 2Ghz processors, requires 3 days. The reconstruction by manifold kernel regressions requires about 30 minutes per image on a 2 Ghz processor.

5 Conclusions

Quantitative evaluation illustrates that the space of brains can be modeled by a low dimensional manifold. The manifold representation of the space of brains can potentially be useful in wide variety of applications. For instance, regression of the parameter space with clinical data, such as MMSE or age, can be used to aid in clinical diagnosis or scientific studies . An open question is whether the manifolds shown here represent the inherent amount of information about shape variability in the data or whether they reflect particular choices in the proposed approach. In particular implementation specific enhancements on image metric, reconstruction, and manifold kernel regression could lead to refined results.

Acknowledgments. This work was supported by the NIH/NCBC grant U54-EB005149, the NSF grant CCF-073222 and the NIBIB grant 5RO1EB007688-02.

References

1. Woods, R.P., Dapretto, M., Sicotte, N.L., Toga, A.W., Mazziotta, J.C.: Creation and use of a talairach-compatible atlas for accurate, automated, nonlinear intersubject registration, and analysis of functional imaging data. Human Brain Mapping 8(2-3), 73–79 (1999)
2. Blezek, D.J., Miller, J.V.: Atlas stratification. Medical Image Analysis 11(5) (2007)
3. Sabuncu, M.R., Balci, S.K., Golland, P.: Discovering modes of an image population through mixture modeling. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 381–389. Springer, Heidelberg (2008)
4. Zöllei, L., Learned-Miller, E.G., Grimson, W.E.L., Wells, W.M.: Efficient population registration of 3D data. In: Liu, Y., Jiang, T.-Z., Zhang, C. (eds.) CVBIA 2005. LNCS, vol. 3765, pp. 291–301. Springer, Heidelberg (2005)
5. Studholme, C., Cardenas, V.: A template free approach to volumetric spatial normalization of brain anatomy. Pattern Recogn. Lett. 25(10), 1191–1202 (2004)
6. Ericsson, A., Aljabar, P., Rueckert, D.: Construction of a patient-specific atlas of the brain: Application to normal aging. In: ISBI, May 2008, pp. 480–483 (2008)
7. Lee, J.A., Verleysen, M.: Nonlinear Dimensionality Reduction. Springer, Heidelberg (2007)
8. Twining, C.J., Marsland, S.: Constructing an atlas for the diffeomorphism group of a compact manifold with boundary, with application to the analysis of image registrations. J. Comput. Appl. Math. 222(2), 411–428 (2008)
9. Davis, B.C., Fletcher, P.T., Bullitt, E., Joshi, S.: Population shape regression from random design data. In: ICCV (2007)
10. Christensen, G.E., Rabbitt, R.D., Miller, M.I.: Deformable Templates Using Large Deformation Kinematics. IEEE Transactions on Medical Imaging 5(10) (1996)
11. Dupuis, P., Grenander, U.: Variational problems on flows of diffeomorphisms for image matching. Q. Appl. Math. LVI(3), 587–600 (1998)
12. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. IJCV 61(2), 139–157 (2005)
13. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290(550), 2319–2323 (2000)
14. Lorenzen, P.J., Davis, B.C., Joshi, S.: Unbiased atlas formation via large deformations metric mapping. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 411–418. Springer, Heidelberg (2005)
15. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23 (2004)
16. Avants, B., Gee, J.C.: Geodesic estimation for large deformation anatomical shape averaging and interpolation. NeuroImage 23(suppl. 1), S139–S150 (2004)
17. Hill, D.L.G., Hajnal, J.V., Rueckert, D., Smith, S.M., Hartkens, T., McLeish, K.: A dynamic brain atlas. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 532–539. Springer, Heidelberg (2002)
18. Zhang, Q., Souvenir, R., Pless, R.: On manifold structure of cardiac mri data: Application to segmentation. In: CVPR 2006, pp. 1092–1098. IEEE, Los Alamitos (2006)
19. Rohde, G., Wang, W., Peng, T., Murphy, R.: Deformation-based nonlinear dimension reduction: Applications to nuclear morphometry, pp. 500–503 (May 2008)
20. Younes, L., Arrate, F., Miller, M.I.: Evolutions equations in computational anatomy. NeuroImage 45(1, suppl. 1), S40–S50 (2009)
21. Hastie, T.: Principal curves and surfaces. Ph.D Dissertation (1984)

Gyral Folding Pattern Analysis via Surface Profiling

Kaiming Li[1,2], Lei Guo[1], Gang Li[1], Jingxin Nie[1], Carlos Faraco[3], Qun Zhao[4], Stephen Miller[3], and Tianming Liu[2]

[1] School of Automation, Northwestern Polytechnical University, Xi'an, China
[2] Department of Computer Science , The University of Georgia, Athens, GA, USA
[3] Department of Psychology, The University of Georgia, Athens, GA, USA
[4] Department of Physics and Astronomy, The University of Georgia, Athens, GA, USA
likaiming@gmail.com

Abstract. Human cortical folding pattern has been studied for decades. This paper proposes a gyrus scale folding pattern analysis technique via cortical surface profiling. Firstly, we sample the cortical surface into 2D profiles and model them using power function. This step provides both the flexibility of representing arbitrary shape by profiling and the compactness of representing shape by parametric modeling. Secondly, based on the estimated model parameters, we extract affine-invariant features on the cortical surface and apply the affinity propagation clustering algorithm to parcellate the cortex into regions with different shape patterns. Finally, a second-round surface profiling is performed on the parcellated cortical regions, and the number of hinges is detected to describe the gyral folding pattern. Experiments demonstrate that our method could successfully classify human gyri into 2-hinge, 3-hinge and 4-hinge gyri. The proposed method has the potential to significantly contribute to automatic segmentation and recognition of cortical gyri.

1 Introduction

The cerebral cortex of human brain is highly convoluted and folds itself into gyri and sulci during brain development. As an essential characteristic of geometry of human cerebral cortex, the folding, however, has shown quite variable patterns on even major gyri and sulci across subjects [1]. Though the mechanisms of cortical folding are still largely unknown [2], evidence has shown that the folding pattern of human cerebral cortex may predict its function [3]. Recently, the quantitative description of folding patterns has emerged as an important research goal [4-6].

The folding pattern of cerebral cortex is a multi-scale concept whose research scope can vary from a very small neighborhood to a whole brain cortical surface (Fig. 1). Currently, there are two major streams of cortical folding pattern analysis. One is based on the very local descriptor curvature and its derivations and whose scope is usually a small neighborhood that is one ring away from the focused vertex (Fig. 1a). In contrast, the other main stream is a quite global one. These latter studies use gyrification index [4] or spherical wavelets [5] to analyze the folding pattern of the whole cortical surface, or at least a certain lobe of human brain (Figs. 1e and 1f). Both techniques have been studied for decades, and have generated many successful applications [7-8]. However, neither of them represents the cortical folding pattern comprehensively and systematically, since

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 313–320, 2009.

essentially cortical folding is a multi-scale concept. One would get quite different descriptions if he/she focused on different scales for the same cortical surface.

This paper proposes a method to analyze the folding pattern of gyri via surface profiling. This is a hybrid parametric method and profiling method in the sense that it combines both advantages of a parametric method (achieving compact representation of shape) and a profiling method (achieving flexibility of arbitrary shape representation). The basic idea is to represent 3D shape information of cortical surface patches with modeling parameters of a series of 2D profiles, and to cluster the cortex into regions with this shape information. Then a second round surface profiling is performed on the gyrus crown of the parcellated cortex, and the number of hinges is detected to describe the folding pattern of the gyrus (please see Fig. 3c for the definitions of a gyrus crown and a hinge). With surface profiling on gyri crowns, we can extend cortical folding analyses from localized parametric representations to gyrus-scale representations.

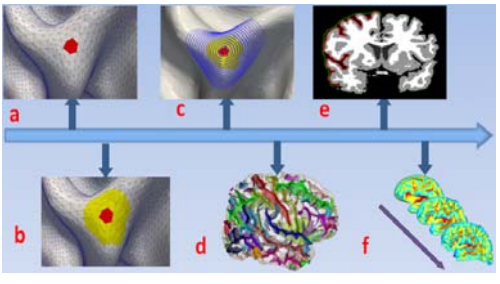


Fig. 1. Multi-scale description of cortical folding patterns. (a): micro-scale (red area, described by curvature); (b): meso-scale (yellow area, described by polynomial model or Bezier surface model); (c): gyrus scale (blue patch, our method); (d): sulcus scale (by [9]); (e): lobe scale (by gyrification index [4]); (f): global scale (by spherical wavelets [5]).

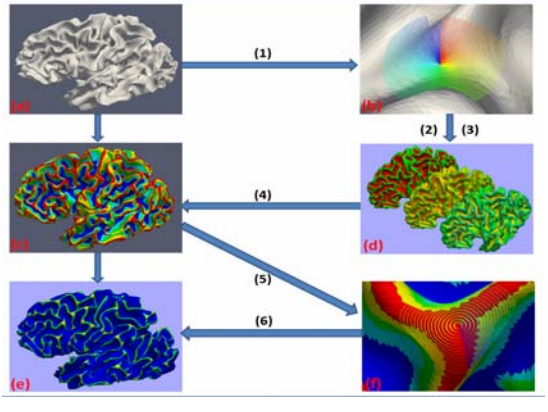


Fig. 2. Flowchart of our surface profiling method. (a): Original cortical surface; (b): Profiles on the original cortical surface; (c): Parcellated cortical surface; (d): Feature surfaces with shape information; (e): Gyral folding pattern surface; (f): Profiles on parcellated cortical surface. (1): Profiling; (2): Model fitting; (3): Feature extraction; (4): Affinity propagation clustering;(5): Profiling; (6): Hinge detection.

2 Method

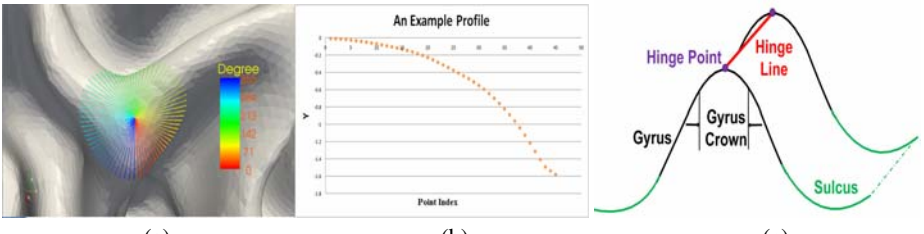
2.1 Overview of the Method

As shown in Fig. 2, our method for gyral folding pattern analysis includes the following steps. Firstly, for each vertex of a reconstructed human brain inner surface, we sample the corresponding patch into a series of 2D profiles, and model these profiles using power function, which is a popular model in structural geology study [12]. The shape information of the current surface patch then is encoded in the parameters of the power function.

Secondly, based on the model parameters and profiling information, we define several affine-invariant features to represent each vertex's folding information, and use these features to cluster the whole cortical surface by the affinity propagation algorithm; this step segments the surface into several major cortical regions, including gyri crowns. Finally, a second round profiling is applied for vertices of gyri crowns on the parcellated cortex, and the number of hinges for each gyrus is detected to represent its folding pattern.

2.2 Profiling of the Cortical Surface

To profile the cortical surface, we first build a 3D coordinate system that combines a 3D Cartesian coordinate system [13] and a 2D polar coordinate system. For any vertex O on the cortical surface, we use its normal direction N as the Z direction in a 3D Cartesian coordinate system, and build a polar coordinate system in its tangent plane P. Then, we start profiling from an arbitrary direction in plane P, and stop profiling for current direction when samples reach a certain maximum M. Profiling is performed every θ degree in plane P. Fig. 3 is an example of profiling.



(a) (b) (c)

Fig. 3. (a) and (b): An example of profiling. (c): Definition of a gyrus crown and a hinge.

The essence of surface profiling here is to simplify a 3D profiling problem down to a collection of 2D profiling problems. This simplification is founded on the fact that the human brain is highly convoluted and surface patches can have very complex shapes. However, current popular models for 3D shapes like polynomials and ellipsoids are symmetric or might be too simple to capture such complex shapes. Thus, the advantage of 2D profiling is the flexibility to describe an arbitrarily shaped cortical surface patch. The disadvantage along with the simplification is the possible loss of 3D shape information. However, the 360 degrees of profiling still captures a great deal of 3D information, especially when we model the profiles and connect corresponding model parameters of all profiles together to form a circle curve.

2.3 Model Fitting

The essential idea of model fitting for profiles is to encode the shape information into a couple of parameters compactly. The model we use in this paper is the power function, a popular model in 2D Geology study because of its simplicity and intrinsic physical meaning [12]. The power function is expressed as:

$$y = y_0 \left(x / x_0 \right)^n \tag{1}$$

Here, x_0, y_0 and n are parameters to describe a profile shape; $y_0 \neq 0, x_0 > 0$ and $n > 0$. The parameters of this model can be evaluated in a sense of least-square using the Levenberg-Marquardt algorithm. Given N sample points of a profile, the parameters are those that minimize the fitting residuals:

$$\hat{P} = \arg\min_{P} \sum_{i=1}^{N} (y_i - y_{pi})^2 \tag{2}$$

Here, P denotes the three parameters to be evaluated, y_{pi} is the model output at the i-*th* point with P given, and y_i is the profile measurement at the i-*th* point.

2.4 Feature Extraction and Clustering

After the model fitting step, we extract several affine-invariant features to represent the shape information of a surface patch. Among the model parameters, the ratio R between y_0 and x_0 and the power n are very information-rich descriptors of the profile shape, especially the ratio R which has proven to be more stable and changes more smoothly between adjacent profiles than the power n. The definitions and descriptions of features that we extract based on model parameters and profiling information are as follows. 1) SulciOrGyri: a vertex that has more profile points above its tangent plane would be considered as on a sulcus. Otherwise, it is on a gyrus. 2) AverRatio: the average R of all profiles for the current vertex. 3) AverMinR: average of Rs that correspond to local minimums at R curve. 4) AverMaxR: average of Rs that correspond to local maximums at R curve. 5) AllDis: sum of distances between neighboring local maximum and minimum. 6) AverDis: average of distances between neighboring local maximum and minimum. 7) MaxDis: maximum of distances between neighboring local maximum and minimum. 8) AverSampleDis: average of the first order moment of all profiles about the tangent plane. 9) MaxSampleDis: maximum of the first order moment of all profiles about the tangent plane. 10) AverPower: average of the power n for all profiles.

Based on the above 10 features, we apply the unsupervised clustering algorithm affinity propagation [14] to the vertices of a cortical surface. The similarity $S(i, j)$ of two random vertices i and j is defined as the minus weighted Mahalanobis distance:

$$S(i, j) = -\sqrt{(\overline{V}_i - \overline{V}_j)^T \, Cov^{-1} W (\overline{V}_i - \overline{V}_j)} \tag{3}$$

Here $\overline{V}_i$ and $\overline{V}_j$ are the feature vector defined above; W is a weighting diagonal matrix with W_{ii} as the weight for the i-*th* feature; Cov is the feature covariance.

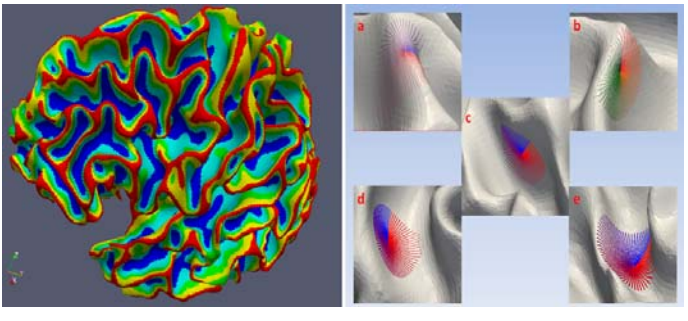


Fig. 4. The five parcellated cortical regions and their typical corresponding patches: gyrus crown (red, a), sub gyrus crown (yellow, b), central area (green, c), sub sulcus basin (light blue, d), and sulcus basin (blue, e)

An example of the clustering result is shown in Fig. 4. The clustering step has two important contributions. First, the cortical surface is automatically parcellated into distinct regions corresponding to gyral and sulcal regions, as well as transitional regions between them. Second, the transition between different cortical regions is smooth. If we walk from gyrus crown to sulcus basin along any path, we will very likely cross the same three transitional cortical regions. These two properties significantly help us profile the gyrus crown on the parcellated cortex and analyze gyral folding patterns.

2.5 Profiling on Gyrus Crown and Hinge Detection

To profile the gyrus crown of the parcellated cortex, we first assign a label to each parcellated cortical region. The assigned labels could reflect the layout and transition between different cortical regions. In this paper, gyri crowns are assigned with a value 1 whereas sulci basins are assigned with a value 5, and other regions are assigned with values according to their transitional levels on the parcellated cortex.

Then, a feature f is created for each profile of gyrus crown in order to measure the profile depth, as well as the number of different regions the profile crosses. The feature is defined as:

$$f = \frac{1}{N}\sum_{i=1}^{N} f_i \tag{4}$$

Here N is the number of points on the profile, and f_i is the label of the region to which point i belongs. For example, if point i is on the gyrus crown, f_i is 1.

Following the above two steps, we detect hinges of the gyrus on which the current vertex sits. After 360 degree profiling, the feature values of all profiles for the current vertex will be combined together to form a ring curve (see Fig. 5d). Local minima of the curve correspond to the hinges of the gyrus, and the number of the local minima is the number of hinges of the gyrus. For example, vertex A in Fig 5 has a very clear folding pattern (three local minima) to indicate that it is on a 3-hinge gyrus.

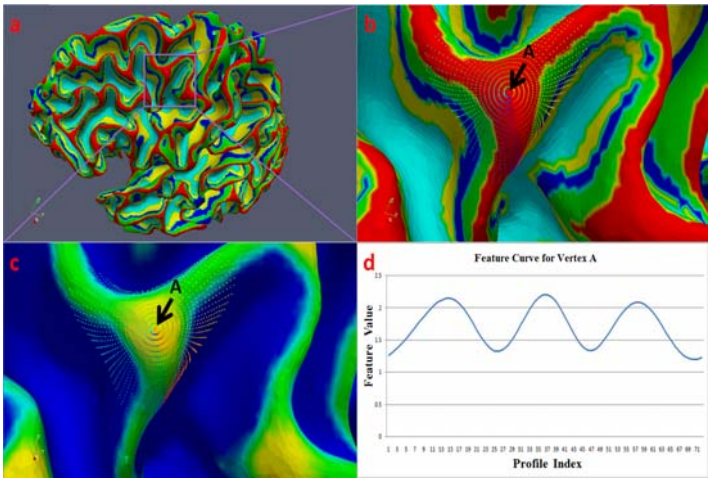


Fig. 5. Profiling of a gyrus crown on parcellated cortex. (a): the parcellated cortex; (b): profiling on a gyrus crown; (c): detected patterns (yellow for 3-hinge gyri, and green for 2-hinge gyri); (d): the feature curve for vertex A.

3 Results

3.1 Gyri Pattern Detection

In this section, we applied the above method to 10 constructed cortical surfaces. Our experimental results demonstrate that human gyral folding patterns can be divided into 3 classes according to their number of hinges: 2-hinge, 3-hinge and 4-hinge gyri. Fig. 6a shows examples of the three folding pattern categories and Fig. 6b shows their corresponding feature curves respectively. As we can see from the figure, the hinges of gyri correspond well to the local minima of the feature curves (Fig. 6b). The number of local minima of the connected feature curve, therefore, is considered as the number of hinges for the current gyrus. Besides the number of local minima of the connected feature curve, the distance between local minima is also an important feature that could be used to further classify the detected gyral folding patterns. This distance actually represents the degree of how much the gyrus bends itself. Take the 2-hinge gyrus in Fig. 6a3 as an example, its bending degree is apparently larger than Fig. 6a1 and Fig. 6a2 in the same category. We can also see the differences from the feature curves in Fig. 6b, that is, the distance of the two local minima in Fig. 6b3 is larger than those of the other two gyri in Fig. 6b1 and Fig. 6b2.

Fig. 6c provides the gyri pattern detection result on a whole cortical surface. Apparently, most of the gyri patterns are correctly detected, indicating reasonably good performance of the proposed method. In particular, the detected 4-hinge patterns are highlighted by dashed circles. One zoomed example is shown in Fig. 6d.

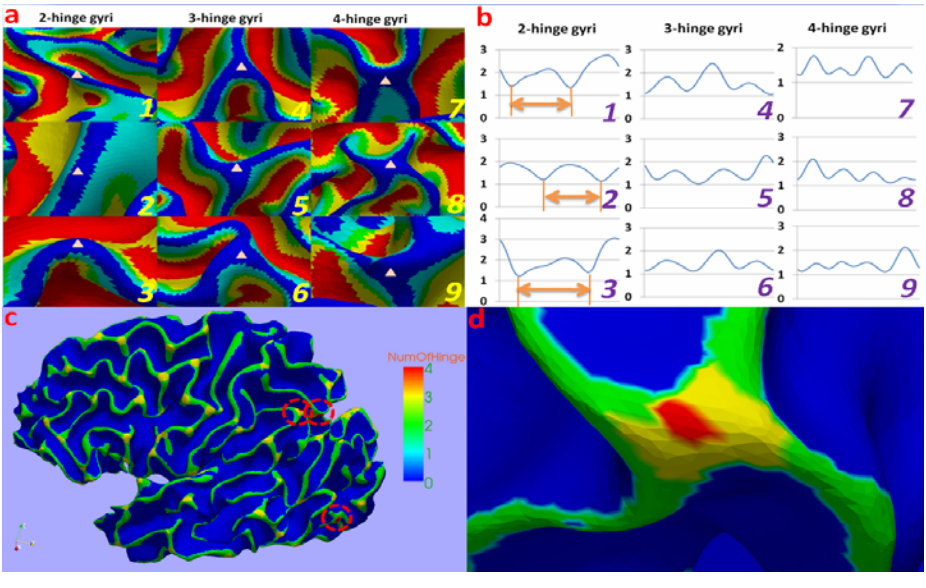


Fig. 6. Pattern detection result. (a): Examples for each detected pattern. Small triangles denote the centers of detected gyri patterns. 1-3: 2-hinge gyri; 4-6: 3-hinge gyri; 7-9: 4-hinge gyri. (b): Corresponding feature curve for each sample. (c): Detected patterns on a whole cortical surface. The three patterns are color-coded. (d): An example of detected 4-hinge gyri.

3.2 Accuracy

To quantitatively evaluate the accuracy of our proposed method, we have two experts manually label the detected patterns, and count the number of two types of detection errors: Type1 error (false positive) and Type2 error (false negative). We express the accuracy as:

$$DetectionAccuracy = (1 - \frac{Type1Errors + Type2Errors}{AllDetectedPatterns}) * 100\% \quad (5)$$

The detection accuracy of the 3-hinge gyri pattern for the randomly selected 10 cortical surfaces is summarized in Table 1. The average accuracy is over 90%. The algorithm has similar accuracy performance for the 2-hinge and 4-hinge gyri pattern detections.

Table 1. Detection accuracy for 3-hinge gyri pattern

Subject ID	Expert 1		Expert 2		Detected Patterns	Accuracy (%)
	Error1	Error2	Error1	Error2		
Sub 1	8	6	7	5	184	92.93
Sub 2	5	9	6	11	196	92.09
Sub 3	6	4	8	3	170	93.82
Sub 4	4	11	4	9	143	90.21
Sub 5	10	9	9	9	168	88.98
Sub 6	6	4	6	5	181	94.19
Sub 7	1	8	3	9	146	92.80
Sub 8	9	3	7	3	203	94.58
Sub 9	10	11	8	10	154	87.34
Sub 10	7	3	5	5	165	93.93
Total	58	70	63	69	1410	90.78

4 Discussion and Conclusion

In this paper, we propose a method to analyze gyral folding pattern via surface profiling. The method focuses on hinge numbers of gyri, and has been applied to 10 normal human brain MR images. Our preliminary results demonstrate that the proposed surface profiling method is able to accurately classify gyri into three folding patterns according to the number of gyral hinges.

In the extant literature, several methods have been proposed to automatically label human brain surface into gyri and sulci [10-11]. In comparison, our segmentation of the cortical surface is based on clustering using profile shape information, and 3 more classes besides gyrus crown and sulcus basin are segmented to fill the transition area from gyri to sulci. Though the segmented gyrus crown might be broken somewhere, it seems that these breaks have little impact on the final results of gyral folding patterns. This robustness may come from the profiling method itself. Since we profile the cortical surface at a macro level, small breaks of a gyrus crown probably would not change the fact that the majority of the profile is on a gyrus.

Our research on gyral folding pattern analysis has shown that 3-hinge and 4-hinge gyri (Fig. 6d) are common across different subjects, and the distribution of them among individuals can vary significantly. This result puts forward new challenges for registration-based analysis of the human brain. For example, how to establish correspondence between different patterns of gyri, e.g., 3-hinge gyri and 4-hinge gyri, in brain registration remains a challenging and open problem.

Currently, our method only classifies gyral folding patterns into 3 broad classes: 2-hinge, 3-hinge, and 4-hinge gyri. A more detailed classification of the folding patterns, however, is possible via surface profiling. For the 2-hinge gyri, we could use the angle between local minima to recognize whether it is a "–" shape gyrus or "U" shape gyrus. For the three-hinge gyri, we could also use the angle information to further classify the gyri into "Y" shapes and "T" shapes. The more detailed classification of 2-hinge gyri and 3-hinge gyri could potentially provide additional important features for self-contained parcellation of the cerebral cortex into anatomically meaningful regions, as well as for automatic recognition of them.

References

1. Talairach, J., Tournoux, P.: Co-planar Stereotaxic Atlas of the Human Brain. Thieme, New York (1988)
2. Van Essen, D.: A tension-based theory of morphogenesis and compact wiring in the central nervous system. Nature 385, 313–318 (1997)
3. Fischl, B., Rajendran, N., Busa, E., et al.: Cortical Folding Patterns and Predicting Cytoarchitecture. Cereb Cortex 18(8), 1973–1980 (2008)
4. Zilles, K., Armstrong, E., Schleicher, A., Kretschmann, H.J.: The human pattern of gyrification in the cerebral cortex. Anat. Embryol. (Berl) 179, 173–179 (1988)
5. Yu, P., Grant, P.E., Qi, Y., Han, X., et al.: Cortical Surface Shape Analysis Based on Spherical Wavelets. IEEE Transaction on Medical Imaging 26(4), 582–597 (2007)
6. Toro, R., Perron, M., Pike, B., et al.: Brain Size and Folding of the Human Cerebral Cortex. Cerebral Cortex 18(10), 2352–2357 (2008)
7. Hardan, A.Y., Jou, R.J., Keshavan, M.S., et al.: Increased frontal cortical folding in autism: a preliminary MRI study. Psychiatry Research: Neuroimaging 131, 263–268 (2004)
8. Rettmann, M.E., Kraut, M.A., Prince, J.L., et al.: Cross-sectional and longitudinal analyses of anatomical sulcal changes associated with aging. Cerebral Cortex 16, 1584–1594 (2006)
9. Mangin, J.F., Rivière, D., Cachia, A., et al.: A framework to study the cortical folding patterns. NeuroImage 23, S129–S138 (2004)
10. Cachia, A., Mangin, J.F., Rivière, D., et al.: A generic framework for parcellation of the cortical surface into gyri using geodesic Voronoi diagrams. Med. Image. Anal. 7(4), 403–416 (2003)
11. Fischl, B., van der Kouwe, A., Destrieux, C., et al.: Automatically parcellating the human cerebral cortex. Cereb. Cortex. 14(1), 11–22 (2004)
12. Bastida, F., Aller, J., Bobillo-Ares, N.C.: Geometrical analysis of folded surfaces using simple functions. Journal of structural geology 21, 729–742 (1999)
13. Tao, X., Prince, J.L., Davatzikos, C., et al.: Using a Statistical Shape Model to Extract Sulcal Curves on the Outer Cortex of the Human Brain. IEEE Trans. Med. Imag. 21(5), 513–524 (2002)
14. Frey, B.J., Dueck, D.: Clustering by Passing Messages Between Data Points. Science 315, 972–976 (2006)

Constrained Data Decomposition and Regression for Analyzing Healthy Aging from Fiber Tract Diffusion Properties*

Sylvain Gouttard[1], Marcel Prastawa[1,2], Elizabeth Bullitt[3], Weili Lin[4], Casey Goodlett[1,2], and Guido Gerig[1,2]

[1] Scientific Computing and Imaging Institute
[2] School of Computing, University of Utah
Departments of [3] Surgery and [4] Radiology, University of North Carolina
gouttard@sci.utah.edu

Abstract. It has been shown that brain structures in normal aging undergo significant changes attributed to neurodevelopmental and neurodegeneration processes as a lifelong, dynamic process. Modeling changes in healthy aging will be necessary to explain differences to neurodegenerative patterns observed in mental illness and neurological disease. Driving application is the analysis of brain white matter properties as a function of age, given a database of diffusion tensor images (DTI) of 86 subjects well-balanced across adulthood. We present a methodology based on constrained PCA (CPCA) for fitting age-related changes of white matter diffusion of fiber tracts. It is shown that CPCA applied to tract functions of diffusion isolates population noise and retains age as a smooth change over time, well represented by the first principal mode. CPCA is therefore applied to a functional data analysis (FDA) problem. Age regression on tract functions reveals a nonlinear trajectory but also age-related changes varying locally along tracts. Four tracts with four different tensor-derived scalar diffusion measures were analyzed, and leave-one-out validation of data compression is shown.

1 Introduction

Unlike earlier hypothesis that brain anatomy and major functions are pretty much established after adolescence and would not change significantly until late adulthood, there is increasing evidence of dynamic, lifelong changes of brain structures and thus plasticity. To examine correlation of age-related brain changes, volumetric measurements such as brain tissue [1] or cortical thickness [2] were proposed. Regression on brain MRI has been used by Davis *et al.*. [3] to

* This work is part of the National Alliance for Medical Image Computing (NA-MIC), funded by the National Institutes of Health through Grant U54 EB005149. We also acknowledge support from the NIMH Silvio Conte Center for Neuroscience of Mental Disorders MH064065, the BRP grant R01 NS055754-01-02 and R01 EB000219 NIBIB.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 321–328, 2009.

depict age-related shape and volume changes. White matter changes in diffusion tensor imaging (DTI) due to aging were reported by analyzing changes of scalar diffusion measures such as fractional anisotropy (FA) and mean diffusivity (MD) in manually selected brain [4,5] or tract regions [6]. Analysis was most often performed by regional measurements of FA or MD across different age groups, or by linear regression of such measurements versus age. Whereas most of this previous analysis uses manually selected regions of interest strategically placed in subregions of known fiber tract locations, clinical research is often interested in the analysis of *whole fiber tracts* associated with specific tasks or cognitive function. Age-related changes of tract diffusion properties should therefore be represented at various positions of tracts, informing researchers about anatomical location and type of diffusion changes. This paper presents a methodology for age regression of fiber tract diffusion properties. We follow the work of Goodlett *et al.*[7] where fiber bundles are parametrized with arc-length and attributed with local diffusion properties summarized across perpendicular cross-sections. These parametrized representations with diffusion attributes can thus represented by functions whose shape represents variation of diffusion as a function of locality, $f(s)$, with $s = [0 \dots l]$. Diffusion is a scalar derived from tensors, so that $f(s)$ can be a function of FA or MD, for example. Measuring such tract functions from subjects with different age t results in a set of functions parametrized as $f(t, s)$. Using a constrained PCA technique (CPCA), we demonstrate how the complex shape and time change information encoded in $f(t, s)$ can be simplified to a model where regression on one coefficient efficiently represents locality, type and magnitude of age-related diffusion changes.

This paper is organized as follows: Section 2 describes the concept of constrained PCA. The processing pipelines for population-based segmentation of fiber tracts and for analysis of age-related diffusion changes of tracts are described in section 3. Section 4 summarizes results on several tracts from the healthy aging study, followed by validation experiments (section 5).

2 Constrained Principal Component Analysis (CPCA)

Our problem is defined as follows: Given a set of parametrized functions of diffusivity attributed with age, $f(t, s)$, we would like to isolate the major systematic change of functional shapes with respect to age from population noise. Such a dimensionality reduction cannot be optimally achieved by principal component analysis (PCA) since any decomposition to modes of major variability would not consider the time attribute. We therefore fit the set of functions into a space that is constrained to a smooth change with time. The first or first few of the resulting coefficients can then be used for age regression via polynomial fit.

We propose the use of constrained Principal Component Analysis (CPCA) [8,9] to perform data reduction that incorporates the age or time effect in the observed data. For a set of discretized functions $f(t, s)$ that represents the observed population data at different points of time, we first construct a data matrix $Z_{t,s}$ indexed by time t along the row and indexed by the tract spatial

parametrization s along the column (as in standard PCA). CPCA linearly separates the data matrix given multiple constraint variables or external information about the structure of the data along the row or columns of the observations. For constraint variables $\alpha_1, \cdots, \alpha_j$, a data matrix M is rewritten as:

$$M = \tilde{M}(\alpha_1) + \cdots + \tilde{M}(\alpha_j) + R \tag{1}$$

where R is the residual data matrix. Unlike standard PCA, CPCA decomposes the individual data matrices for each constraint variable rather than the original unconstrained data matrix. This ensures that the subspaces generated by CPCA encode the effects of the constraint variables. In our tract analysis, we have one constraint variable along the row (time) and Z can be formulated as:

$$Z = \tilde{Z}(t) + R \tag{2}$$

where $\tilde{Z}(t)$ is a projection of each individual function onto the space of polynomials $a_0 + a_1 t + \cdots + a_k t^k$ that smoothly varies with time and R is the residual data matrix. We have chosen the projection to the space of polynomials to constrain the analysis to changes that vary smoothly with time and exclude noise and population variability. Given a matrix of polynomial basis functions B and the matrix of polynomial coefficients that best fit the observed data C, $\tilde{Z}(t) = BC$ and $Z = BC + R$.

Within the constrained CPCA framework, all analysis of the diffusion properties is performed within the subspaces of the constrained data matrix BC and we exclude the residual matrix to ensure that noise is not attributed as age effect. A polynomial up to a degree 4 was used in our experiments.

3 Application to Age-Related White Matter Tract Changes

Driving Application: White Matter Change in Healthy Aging. We have access to a database of high-resolution, 3-Tesla MR images of 100 healthy subjects aged 20 to 76, with 20 subjects per decade divided equally by gender. All subjects are carefully screened for the presence of brain-related disease. Images acquired include T1, T2, MRA, and diffusion tensor images. After screening for image quality of successful DTI scans, 86 DTI could be selected for this study. DTI was done on a Siemens Allegra 3T head-only scanner: One image without diffusion gradients together with diffusion weighted images along six gradient directions with a b-value of 850, repeated 2 times for averaging. TR/TE were 7500/73ms, the matrix size was 64x64, and voxel resolution was isotropic with $2x2x2mm^3$.

Atlas Building of Tensor Images. The images are initially aligned by affine registration of the non-diffusion weighted images to a template T2 (MNI) image using mutual information, which ensures mapping into a normative brain space. A scalar feature, the greatest eigenvalue of the Hessian of the FA image, is computed for each subject [7]. A diffeomorphic, fluid registration is computed for the set of feature images [10] using the affine alignment as an initalization.

The result of the registration is a set of transformations that smoothly map each subject image into a common atlas coordinate system and can also be used to map structures from the atlas back to each subject's native space. All tensor images deformed into the atlas space are averaged to produce a template tensor image for the atlas, which is used to extract fiber tracts for the white matter structures of interest. We extracted the four bundles genu, splenium, mid corpus callosum (mid-cc) and motor-sensory via fiber tractography in atlas space to ensure correspondence across subjects (Fig. 1 right), using a standard streamline integration method. These tracts were then mapped back into each subjects native space to obtain diffusion measures for that subject. This results in a set of tracts in the atlas space with equivalent geometry but diffusion values extracted from each subject. Validation and comparison to individual tractography has been done, but space constraints prohibit inclusion in this paper. Since subjects have different ages equally distributed between 19 and 68, we therefore get sets of functions represented as $f(t, s)$ with s representing parametrization by arc-length and t subject age. The scalar diffusion measures f can be chosen as FA, MD (trace), axial diffusivity (AD, λ_1) and radial diffusivity (RD, $(\lambda_2 + \lambda_3)/2$).

CPCA on a Set of Functions. Figure 2 illustrates the example of FA of the mid-cc tract represented by a set of functions coded with age. We can observe that FA at the middle part close to the corpus callosum does not change strongly with age but that the sidelobes, i.e. bundle locations close to the cortex, show significant decrease of FA with age. The complexity of age change and the type and amount of change at different tract locations are not visible due to large population variability. We then applied constrained CPCA to the set $FA(t, s)$ in order to decrease dimensionality and determine major shape variation of the FA functions with age. Standard PCA is applied for comparison.

Figure 3 illustrates results for PCA (top) and CPCA (bottom) for the mid-cc tract. Plots in the left column show samples projected into the space of the first two principal components and age. Samples are color-coded with age. As expected, PCA does not present obvious structure versus age but CPCA results in a low-dimensional continuous path for the 86 samples. The middle column illustrates projection to only the first component versus age, with overlay of 4th-order regression on age. Again, PCA does not depict a clear relationship to age, but age regression of CPCA (middle, bottom) illustrates a nonlinear continuous change with age, with large changes between age 20 and 30, than flattening till age 50, and accelerated change thereafter. The right column displays the FA tract functions reconstructed from the first component for all 86 subjects. Unlike PCA, CPCA encodes age change of the FA functions, clearly visible as the systematic age-sorted coloring of CPCA and random coloring of PCA.

Our main aim is a simplified model of changes of diffusion functions with respect to age. As seen in Fig. 4(b) left for the mid-cc tract, most energy is concentrated on the first component of CPCA. Moreover, age correlation is very strong for the first component but nonsignificant for the others. PCA is shown for comparison. We therefore chose to model the diffusion changes by age regression within the subspace of only the first component. Here, we fit a polynomial up to

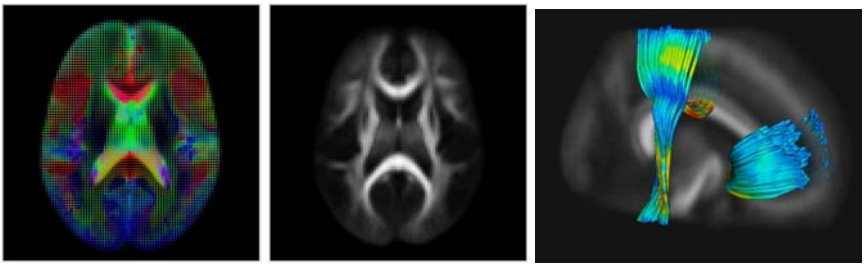


Fig. 1. DTI population atlas: Left Tensor field of atlas computed over 86 subjects: Middle: FA map calculated from tensor atlas. Right: Typical fiber bundles used in this study. Mid corpus callosum, sensory-motor and genu tracts are color-coded with FA, with $[0 \cdots 1]$ ranging from dark blue over yellow to red.

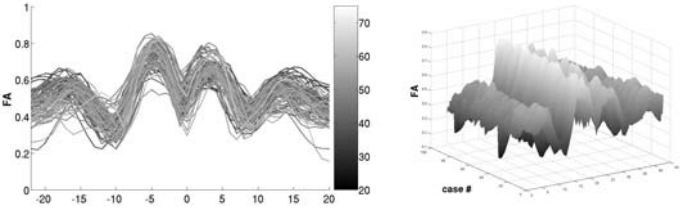
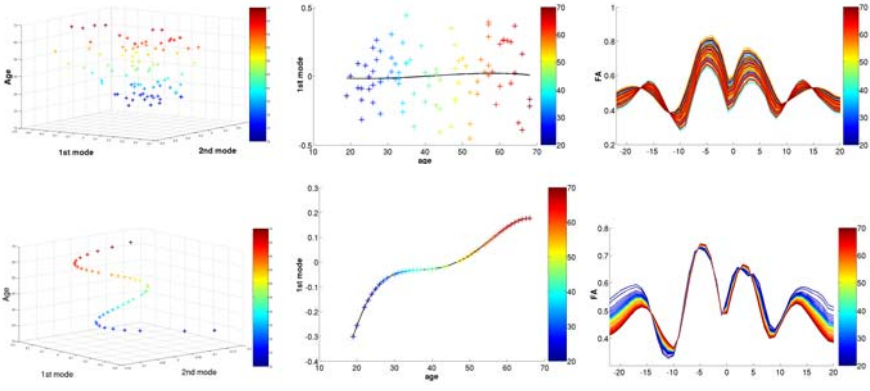
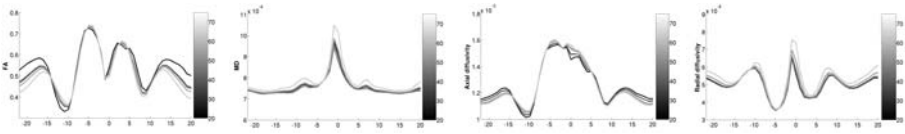


Fig. 2. Plot of all mid-cc tract functions of FA shaded by age (left) and 3D representation of the age set sorted by age (young to old from right to left)



a b c

Fig. 3. Results of the analysis of FA of the mid-cc tract with standard PCA (top) and constrained CPCA (bottom). Age is encoded in color. a) The first two modes plotted against age. b) Only first mode versus age with overlay of 4th order age regression. c) FA functions of 86 subjects projected onto the first component. Columns b and c demonstrate that data reduction by CPCA does encode a strong age-related relationship not seen in PCA (see random coloring in PCA versus age-sorted coloring in CPCA in column c).



FA MD AD RD

(a) Reconstruction of diffusion properties of the mid-cc tract using polynomial age fitting within the first subspace generated by CPCA. Age is encoded in different shades of gray amd shown for years 20,30,40,50 and 60. We display the evolution of diffusion properties as a function of age for FA, MD, AD and RD.

	CPCA		PCA	
Mode	%eval	corr	%eval	corr
1	66.20	**0.978**	24.25	0.069
2	23.68	0.032	20.82	-0.496
3	7.78	0.020	12.86	0.145
4	2.33	0.203	10.81	-0.047

	CPCA %eval			
Mode	FA	MD	AD	RD
1	66.20	80.98	55.40	64.91
2	23.68	15.91	21.92	25.93
3	7.78	2.54	19.70	6.87
4	2.33	0.56	2.96	2.27

(b) Left: Normalized eigenvalues and age correlation of first four CPCA and PCA modes on the set of functions of FA along the mid-cc tract, illustrating the stronger compression of CPCA but also concentration of age-correlation on the first component. Right: Normalized eigenvalues of CPCA for FA, MD, AD and RD diffusion measures.

Fig. 4. Results of mid-cc tract for four different diffusion measures

a degree 4 to the projections of the functions to the first component (see Fig. 3 column b). Using this polynomial fit, we can then reconstruct the expected diffusion properties for each age in the range of interest. Fig. 4(a) illustrates the reconstructions for ages 20,30,40,50 and 60 and for various diffusion properties (FA, MD, AD, RD) along the mid-cc tract. Age is encoded using a gray-scale color map. The figures show nonlinear differences across age and also variations of change patterns as a function of tract location.

4 Results

The CPCA analysis has been applied to the major four fiber tracts genu, splenium, mid-cc and motor-sensory. Fig. 5 displays age-related changes as calculated for these four tracts. We illustrate the reconstructed diffusion functions of FA and MD for age 20, 30, 40, 50 and 60, using projection onto the first CPCA component and 4th-order regression. The plots clearly illustrate that the age-related trajectory of white matter diffusion is nonlinear, i.e. different decades present different changes, and second that diffusion changes vary significantly as a function of tract location. FA in genu, splenium and mid corpus callosum, for example, shows only minor age change in the center region but significant FA decrease in the peripheral parts close to the cortex. The motor-sensory tract (arc-length left to right presents superior to inferior tract location, see also Fig. 1) also demonstrates age-changes of FA close to the cortex but large MD changes along the middle part between cortex and internal capsule. The splenium tract showed much more population variability than the genu, which is reflected in

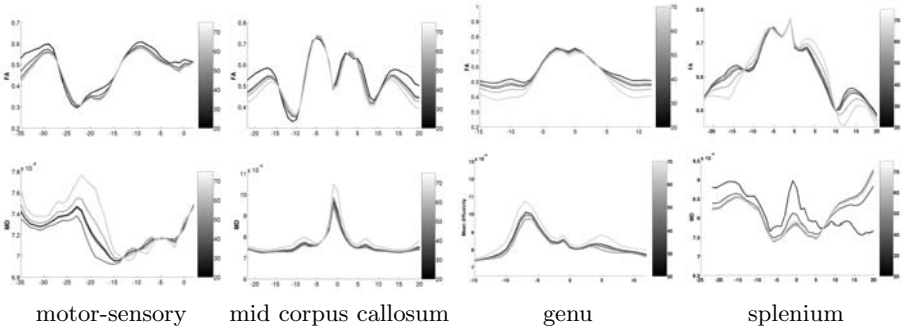


Fig. 5. Reconstructed functions of FA (top row) and MD (bottom row) for four fiber tracts. Age is encoded as gray-scale for years 20, 30, 40, 50, and 60. Age-correlation of CPCA was significant for all 4 tracts (0.93, 0.98, 0.85 and 0.95).

the MD analysis. Major conclusions related to the specific age study might be too preliminary, but this discussion demonstrates how the results could be interpreted by clinical researchers. The caption in Fig. 5 also lists the age correlation for the first component of CPCA which is highly significant for all 4 tracts.

5 Validation

We validate the choice of an age-constrained subspace by leave-one-out cross validation. Given a set of N functions, we perform leave-one-out analysis by excluding one function from the set followed by performing data reduction and regression using the reduced set. We then project the one excluded function to the computed first principal component, and compare the projected score against the expected score from the polynomial fitting. We decided to use this approach to validate the data reduction by projection to a lower dimensional subspace and also the age-based fitting within this subspace. This analysis was performed on the subspaces computed by CPCA and regular PCA. The values for the average of the $sum_x|proj(x) - fit(x)|^2$ measure for motor-sensory, mid-cc, genu and splenium are (0.01975, 0.02298, 0.01698, 0.03778) for CPCA and (0.02499, 0.03051, 0.02367, 0.12504) for PCA. As seen, the CPCA consistently provides lower prediction error compared to the unconstrained PCA. The prediction error for PCA is higher for the noisy diffusion data in the splenium tract.

6 Discussion

We present a methodology for analyzing age-related changes of white matter measured along fiber tracts. The framework includes population-based mapping of DTI to a common coordinate space, tractography, and representation of tensor-derived scalar diffusion measures along tracts. CPCA constrained to age is then applied to the sets of diffusion functions to reduce dimensionality to the

first principal mode. The PCA projections are performed on the curves reconstructed from the polynomials. We thus enforce the dominant eigenvectors to be strongly driven by time and not necessarily be the original eigenvectors. We preferred this global approach versus regression at each spatial position to account for along-tract correlation. Results on major tracts demonstrate feasibility and represent type of changes of FA and MD as reported elsewhere. However, our preliminary analysis predicts that age changes might not be simplified to linear regression but show more complex time and space variations. The strong dimensionality reduction while still encoding age-related changes as a function of tract location might be a benefit for biostatisical analysis. We plan to apply the method to clinical studies of aging and early brain development, with testing for group differences of change trajectories of white matter diffusion. A limitation might be the independent analysis of tensor-derived diffusion parameters. In the future, we will explore joint analysis of multiple features, e.g. FA and tensor norm or AD and RD, or even full tensors.

References

1. Mortamet, B., Zeng, D., Gerig, G., Prastawa, M., Bullitt, E.: Effects of Healthy Aging Measured By Intracranial Compartment Volumes Using a Designed MR Brain Database. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 383–391. Springer, Heidelberg (2005)
2. Sowell, E.R., Thompson, P.M., Leonard, C.M., Welcome, S.E., Kan, E., Toga, A.W.: Longitudinal mapping of cortical thickness and brain growth in normal children. J. Neurosci. 24, 8223–8231 (2004)
3. Davis, B., Fletcher, P., Bullitt, E., Joshi, S.: Population Shape Regression From Random Design Data. In: Int. Conf. Comp. Vis., ICCV 2007, pp. 1–7 (2007)
4. Salat, D., Tuch, D., Greve, D., van der Kouwe, A., Hevelone, N., Zaleta, A., Rosen, B., Fischl, B., Corkin, S., Rosas, H., et al.: Age-related alterations in white matter microstructure measured by diffusion tensor imaging. Neurobiology of Aging 26(8), 1215–1227 (2005)
5. Pfefferbaum, A., Sullivan, E.: Increased brain white matter diffusivity in normal adult aging: Relationship to anisotropy and partial voluming. Magnetic Resonance in Medicine 49(5), 953–961 (2003)
6. Bonekamp, D., Nagae, L., Degaonkar, M., Matson, M., Abdalla, W., Barker, P., Mori, S., Horská, A.: Diffusion tensor imaging in children and adolescents: Reproducibility, hemispheric, and age-related differences. Neuroimage 34(2), 733–742 (2007)
7. Goodlett, C.B., Fletcher, P.T., Gilmore, J.H., Gerig, G.: Group analysis of DTI fiber tract statistics with application to neurodevelopment. NeuroImage 45(1, suppl. 1), S133–S142 (2009); Mathematics in Brain Imaging
8. Takane, Y., Hunter, M.A.: Constrained principal component analysis: a comprehensive theory. Applicable Algebra in Engineering, Communication, and Computing 12, 391–419 (2001)
9. Hunter, M.A., Takane, Y.: Constrained principal component analysis: various applications. Journal of Educational and Behavioral Statistics 27, 41–81 (2002)
10. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)

Two-Compartment Models of the Diffusion MR Signal in Brain White Matter

Eleftheria Panagiotaki[1], Hubert Fonteijn[1], Bernard Siow[1,2], Matt G. Hall[1], Anthony Price[2], Mark F. Lythgoe[2], and Daniel C. Alexander[1,⋆]

[1] Centre for Medical Image Computing, Department of Computer Science, University College London, UK
E.Panagiotaki@cs.ucl.ac.uk
[2] Centre for Advanced Biomedical Imaging , University College London, UK

Abstract. This study aims to identify the minimum requirements for an accurate model of the diffusion MR signal in white matter of the brain. We construct a hierarchy of two-compartment models of white matter from combinations of simple models for the intra and extra-cellular spaces. We devise a new diffusion MRI protocol that provides measurements with a wide range of parameters for diffusion sensitization both parallel and perpendicular to white matter fibres. We use the protocol to acquire data from a fixed rat brain, which allows us to fit, study and compare the different models. The results show that models which incorporate pore size describe the measurements most accurately. The best fit comes from combining a full diffusion tensor (DT) model of the extra-cellular space with a cylindrical intra-cellular component.

1 Introduction

Over the last 15 years diffusion-weighted MRI (DW-MRI) has become popular because it provides unique insight into brain tissue microstructure and connectivity. The technique has become an essential probe for highlighting and monitoring tissue microstructure changes in development and disease.

The simplest and most commonly used model for relating the DW-MRI signal to diffusion in tissue is the diffusion tensor (DT) [1]. The model provides useful microstructural markers of tissue integrity such as fractional anisotropy (FA) and mean diffusivity (MD). However these indicators are non-specific, because many features of the microstructure can affect them. To address this limitation, the recent trend in diffusion MRI [2,3,4] is towards more direct microstructure imaging via more descriptive models of tissue that relate specific parameters, such as cell size and density, directly to the signal.

To trust the parameter estimates we obtain from fitting these models, we must ensure that they include all the important physiological parameters of the tissue

⋆ Thanks to Mankin Choy and Johannes Riegler for help with the sample preparation. Funded by EPSRC grants EP/E056938/1 and EP/E007748, British Heart Foundation and BBSRC.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 329–336, 2009.

that affect the signal. In this work, we study one important class of models: two-compartment models with no exchange, which model the signal as the sum of signals from water inside and outside impermeable cells. The class of models includes the ball and stick model [2], CHARMED [4] and the simplified version of CHARMED in [3], each of which uses different intra-cellular (IC) and extra-cellular (EC) models. We construct a hierarchy of two-compartment models from combinations of these IC and EC models. We define a new diffusion MRI protocol to allow evaluation and comparison of the models for parallel and perpendicular signals in brain white matter over a wide range of scan parameters. Such a study is challenging in brain tissue because fibre orientation varies, so most previous studies use simpler tissue samples such as spinal cord. The new protocol enables extension of these studies to the brain. Here we acquire data from an ex vivo rat brain that has been perfusion fixed. The data set contains a much more comprehensive set of measurements than we can acquire on live subjects but is rich enough to ensure good fit of the models, identification of important effects and thus reliable selection of appropriate models for sparser in vivo data.

Section 2 gives some background on diffusion MRI models. Section 3 introduces the hierarchy of models to test, the MRI acquisition protocol and the model fitting procedure. Section 4 describes the experiments and results.

2 Background

Diffusion tensor imaging (DTI) [1] models the displacement of spins in 3D with a zero-mean Gaussian distribution by fitting the apparent diffusion tensor $\mathbf{D}$, to six or more normalised DW images via

$$S(\mathbf{G}, \Delta, \delta) = S_0 \exp(-b\hat{\mathbf{G}}^T \mathbf{D} \hat{\mathbf{G}}), \tag{1}$$

where S is the DW signal, S_0 is the unweighted signal, b is the diffusion weighting factor, equal to $(\Delta - \delta/3)(\gamma\delta|\mathbf{G}|)^2$ for the pulse-gradient spin-echo (PGSE) sequence , $\mathbf{G}$ is the gradient vector with strength $|\mathbf{G}|$ and direction $\hat{\mathbf{G}}$, Δ is the time between the onsets of the two pulses, δ is the pulse gradient duration and γ is the gyromagnetic ratio. A big limitation of the DT model is that it does not account for restricted diffusion within cells so the signal departs from the model even in single fibre populations especially as b becomes large.

Two-compartment models overcome the limitations of the DT model to some extent by modelling restriction within cells. Often they assume hindered diffusion in the EC space and restricted diffusion in the IC space. The signal attenuation is given by

$$S(\mathbf{G}, \Delta, \delta) = S_0 \Big(f S_h(\mathbf{G}, \Delta, \delta) + (1 - f) S_r(\mathbf{G}, \Delta, \delta) \Big), \tag{2}$$

where f is the volume fraction of the restricted IC compartment, S_h and S_r are signals from the hindered and the restricted compartments respectively.

In Behrens' ball and stick model [2] both S_r and S_h have the form of Eq. 1. For S_r, $\mathbf{D} = d\mathbf{n}\mathbf{n}^T$ where d is the free diffusivity and $\mathbf{n}$ is the fibre direction, so water moves only in the fibre direction. For S_h, $\mathbf{D} = d\mathbf{I}$, where $\mathbf{I}$ is the identity tensor, so diffusion is isotropic.

Alexander [3] models the EC space using Eq. 1 with a cylindrically symmetric tensor, so $\mathbf{D} = \alpha\mathbf{n}\mathbf{n}^T + \beta\mathbf{I}$. The model for the IC space accounts for non-zero pore size, unlike Behrens' stick model. S_r comes from Van Gelderen's model [5] for signal attenuation from diffusion perpendicular to the axis of a restricting cylinder and accounts for finite pulse width.

Assaf's CHARMED model [4] also assumes cylindrical restriction in the IC space. They use Neuman's expression [6] for diffusion in cylindrical confinement for a PGSE experiment, which assumes short gradient pulses ($\delta \ll \Delta$). Unlike Alexander's model, which assumes a single cylinder radius, the model assumes cylinders with gamma distributed radii which introduces one extra parameter. The hindered compartment uses the full DT model constrained only to have principal direction aligned with the cylinder axis.

Other methods describe diffusion with three or more compartments and allow exchange between them. For example, Stanisz et al. [7] construct a three-pool model with prolate ellipsoidal axons and spherical glial cells each with partially permeable membranes. However, fitting such models requires very high quality measurements, typically using NMR spectroscopy rather than MRI. We limit investigation here to simpler two-compartment models with no exchange.

3 Methods

This section describes the hierarchy of two-compartment models, the details of the new acquisition protocol and the model fitting procedure.

3.1 Model Hierarchy

There are many options in the literature for models of the S_r and S_h components from which we can combine any pair to create new two-compartment models.

We investigate two models for the restricted compartment. The first is Behrens' "stick" model [2] which has $\mathbf{n}$ and d as parameters. This describes diffusion in a zero radius cylinder. The second is Van Gelderen's model [5], as used in [3], which accounts for non-zero cylinder radius. We refer to this model as the "cylinder". This model has an extra parameter, R, the axon radius. We assume a single R as in [3].

We investigate three models for the EC compartment. Each is a DT model with different constraints. The first, the "ball", is isotropic, $\mathbf{D} = d\mathbf{I}$, as in [2] and has only one parameter, the diffusivity d. The second has a DT which is anisotropic, but cylindrically symmetric, as in [3] and we call this a "zeppelin". The model has parameters: $\mathbf{n}$, $d_\parallel$ which is the diffusivity parallel to the fibre direction and $d_\perp$ perpendicular to it.

Finally we consider a full tensor. This model has three diffusivity parameters: parallel diffusivity $d_\parallel$ and $d_{\perp_1}$,$d_{\perp_2}$ perpendicular with $d_{\perp_1} \neq d_{\perp_2}$. It has an additional three degrees of freedom for the orthogonal eigenvectors $\mathbf{n}$, $\mathbf{n}_{\perp_1}$ and $\mathbf{n}_{\perp_2}$. The form of the DT is

$$\mathbf{D}_h = d_\parallel \mathbf{n}\mathbf{n}^T + d_{\perp_1}\mathbf{n}_{\perp_1}\mathbf{n}_{\perp_1}^T + d_{\perp_2}\mathbf{n}_{\perp_2}\mathbf{n}_{\perp_2}^T. \quad (3)$$

We refer to this EC model as a "boat". We use combined terms to refer to specific two-compartment models, for example "zeppelin and stick" assumes zero radius cylinders for the IC space and cylindrical symmetry for the EC space. Where appropriate we further constrain the EC models so that the DT's principal direction is equal to the IC cylinder axis and the EC diffusivity is $d_{\|}$.

3.2 MRI Acquisition

We are interested in diffusion in the brain in directions parallel and perpendicular to the fibre orientation, since these directions reveal most about the underlying brain microstructure. However, it is challenging to acquire such measurements from brain tissue, because fibre orientation varies throughout the white matter. Our approach is to pick one central parallel direction and several perpendicular directions and identify voxels in which the fibres align with those directions after imaging. We focus analysis only on these voxels and discard all others.

We acquire diffusion-weighted MR images of a perfusion-fixed male rat brain, using a small bore 9.4T scanner (Varian) with maximum gradient strength 400 mT/m. We use a five direction-encoding scheme and place the sample in the scanner oriented to ensure that some fibres in the Corpus Callosum (CC) are parallel to our central direction. We choose four evenly spaced directions perpendicular to the central direction in our chosen voxels (see Figure 1). We use a PGSE sequence for 70 diffusion weightings: five diffusion times Δ = 10, 20, 30, 40, 50 ms, gradient durations δ = 3 ms for all Δ and δ = 30 ms for Δ = 40, 50 ms, gradient strength $|\mathbf{G}|$ varied from 40 to 400 mT/m in ten steps of 40 mT/m. Measurements with b value greater than 7.7×10^{10} sm^{-2} were not performed due to poor SNR (< 2.6 i.e all combinations with $|\mathbf{G}| = 200$ to 400 mT/m and $\delta = 30$ ms). In total we acquired images with 59 diffusion weightings in each direction. In figure 2 we plot the parallel and the mean of the four perpendicular directions of the log signal from voxels in the region of interest (see section 4) in the CC and demonstrate MRI images for various b values.

We use minimum echo times (TE) to maximise SNR and repetition times (TR) to minimise gradient heating effects. For each combination of diffusion weighting parameters we acquire $b = 0$ images to correct for T1 and T2 dependence. We also perform a separate DTI acquisition using a 42-direction scheme with b value 4.5×10^9 sm^{-2} and six $b = 0$ measurements. The in-plane field of view is 2 cm, matrix size 256×256 and slice thickness 0.5 mm.

3.3 Model Fitting

We fit each model to the data using an iterative optimization procedure and synthesise diffusion-weighted data from the fitted models. We minimize the sum of squared errors using a Levenberg-Marquardt algorithm. Fits of the simplest models are relatively independent of starting position. More complex models are more sensitive and we use parameter estimates of simpler models to provide initial estimates. We choose the best fit parameters from the models after 500 perturbations of the starting parameters to ensure a good minimum. Fitting procedures are implemented in Camino [8].

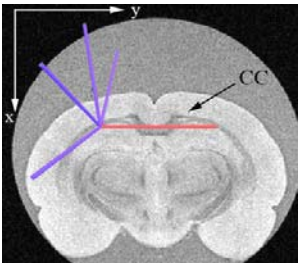


Fig. 1. The red arrow indicates the central gradient direction used for the encoding scheme and the blue arrows indicate the four directions perpendicular to the central one

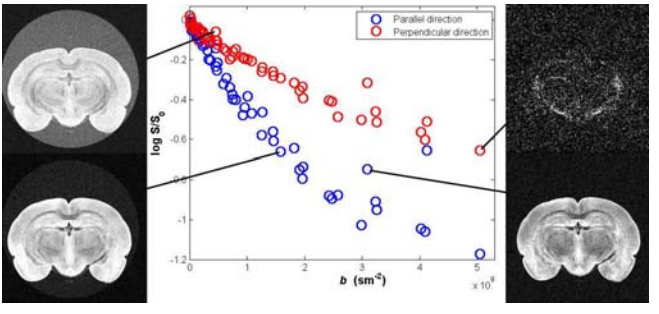


Fig. 2. Plot of the parallel and the mean of the four perpendicular directions of the log signal from voxels in a region of interest in the CC and demonstration of MRI images for various b values

4 Experiments and Results

To study parallel and perpendicular signal attenuation we choose a region of interest (ROI) with fibre direction aligned to the central direction. We manually segment the CC on a FA map from the DTI acquisition and threshold for voxels with FA > 0.5 in which the principal direction of the DT is parallel to the central gradient direction within $2\,^{\circ}$. We average the data contained within all the resulting 21 voxels of the ROI.

The best fit microstructure parameters from the models are shown in table 1. We see that the "cylinder" models give higher values for the volume fraction f and diffusivity parameters and consistently estimate R around $2\,\mu$m, which is a reasonable estimate of mean axon radius. All the models give good estimates of the left-right fibre direction.

The Bayesian information criterion (BIC) [9] evaluates the models and accounts for varying complexity. Table 2 shows the mean-squared error (MSE) and BIC for all the models. As expected, MSE decreases with model complexity, but the BIC reveals which reductions are significant. The "boat and cylinder" model minimizes the BIC.

Figure 3 shows the fit of each model to the data. The top right panel shows scan data and the other panels compare predictions from each model with fitted parameters. The plots actually show the mean signals over 500 trials adding independent Rician noise at approximately the level in the scan data. This procedure

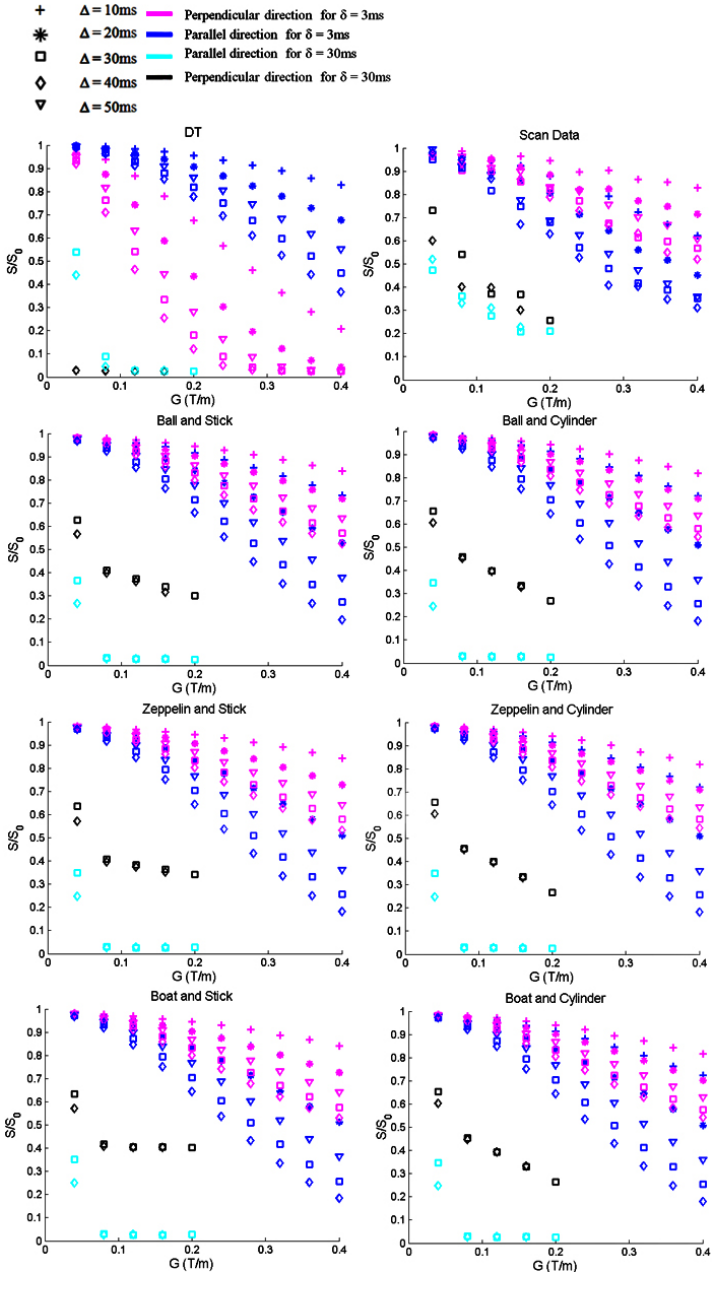


Fig. 3. Results of data synthesised from the analytical models and the scan data from the PGSE experiment. The normalised signal S/S_0 is plotted for all the values of Δ, δ as a function of the gradient strength $|\mathbf{G}|$ for the parallel and the mean of the four perpendicular directions.

only significantly affects the measurements with very low signal. We compare data synthesised from the models by plotting the normalised signal S/S_0 at all values of Δ and δ as a function of the gradient strength $\mathbf{G}$ for the parallel and perpendicular directions. The DT model shows a significant departure from the scan data and confirms expectations that the model is poor for high b value data, because it does not account for restriction. In contrast, all the two-compartment models capture the broad trends of the data and the anisotropy that separates the parallel and perpendicular signals. The subtle variations that improve the fit for cylinder and anisotropic models are difficult to observe qualitatively.

Table 1. Fitted parameters for each model

Models	f	$d_{\parallel}$ ($\mathrm{m^2 s^{-1}}$)	$d_{\perp_1}$ ($\mathrm{m^2 s^{-1}}$)	$d_{\perp_2}$ ($\mathrm{m^2 s^{-1}}$)	R (m)	θ	φ
DT	n/a	2.494×10^{-11}	1.781×10^{-11}	1.660×10^{-11}	n/a	1.570	4.712
Ball and stick	0.429	3.211×10^{-10}	n/a	n/a	n/a	1.570	4.712
Ball and cylinder	0.503	3.389×10^{-10}	n/a	n/a	2.058×10^{-6}	1.518	4.753
Zeppelin and stick	0.408	3.3676×10^{-10}	2.914×10^{-10}	n/a	n/a	1.491	4.790
Zeppelin and cylinder	0.503	3.387×10^{-10}	3.396×10^{-10}	n/a	2.059×10^{-6}	1.518	4.754
Boat and stick	0.410	3.363×10^{-10}	3.772×10^{-10}	2.179×10^{-10}	n/a	1.682	4.704
Boat and cylinder	0.499	3.381×10^{-10}	4.442×10^{-10}	2.405×10^{-10}	2.101×10^{-6}	1.579	4.707

Models	MSE $\times10^4$	BIC $\times10^{-3}$	No. parameters
DT	1025.4	-0.7651	7
Ball and stick	34.5	-1.9768	5
Ball and cylinder	30.8	-2.0116	6
Zeppelin and stick	33.3	-1.9832	6
Zeppelin and cylinder	30.8	-2.0058	7
Boat and stick	30.3	-2.0051	8
Boat and cylinder	27.4	-2.0349	9

Table 2. Mean-squared fitting error, Bayesian information criteria and the number of parameters (with S_0) for each model

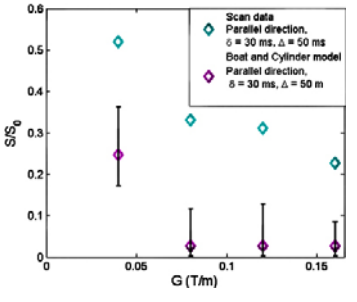


Fig. 4. Plot of the normalised signal for the parallel direction at $\delta = 30\,\mathrm{ms}$ and $\Delta = 50\,\mathrm{ms}$ from the scan data and the "boat and cylinder" model against the gradient strength. The error bars indicate the minimum and maximum signal over 500 Rician noise trials

The biggest departures are for large δ in the parallel direction. We hypothesise that these departures are not solely due to noise. Figure 4 compares the normalised signal S/S_0 for the scan data and the "boat and cylinder" model with $\delta = 30\,\mathrm{ms}$ and $\Delta = 50\,\mathrm{ms}$ for the parallel direction indicating the range of Rician noise over 500 realisations, confirming this.

5 Conclusions

We have constructed, evaluated and compared a hierarchy of two-compartment models for the DW-MRI signal in white matter. Previous studies of this type [4,7] have been limited to non-brain tissue, such as spinal cord or nerve tissue samples, because of the difficulty of obtaining parallel and perpendicular measurements

consistently from the brain. The new imaging protocol we devise provides parallel and perpendicular signals from brain tissue.

The models we present here are applicable and feasible for whole-brain imaging but the acquisition protocol is purposefully not. The aim of this work is to compare models with fixed orientation, which allows many more measurements to support the model comparison. Once we have established appropriate models, subsequently we can find more economical protocols for whole-brain imaging, using for example the ideas in [3].

The key conclusion is that the effects of restriction are extremely important for modelling diffusion in white matter. The simple two-compartment models we study here explains the data remarkably well. Even the simplest four-parameter (excluding S_0) ball and stick model captures the broad trends in the data where the six-parameter DT model completely fails. The model comparison in table 2 clearly demonstrates that the data supports the non-zero axon radius parameter, adding credence to techniques that estimate the parameter [3,4,7], as well as anisotropy of the EC compartment. The departure of the signals from the model in the parallel direction most likely comes from a small amount of restriction parallel to the fibres from glial cells and/or non-parallel fibres, which supports Stanisz's findings in [7], or from effects of tissue fixation.

Future work will extend the hierarchy to other IC and EC models and compare models that include exchange between compartments and different diffusivity and relaxivity in each compartment, as in [7]. In particular, we could include the gamma distribution model of the cylinder radii in [4]. The imaging protocol and methods we use here extend easily to support comparisons of an extended hierarchy of models.

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biop J. 66, 259–267 (1994)
2. Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen, H.: Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077–1088 (2003)
3. Alexander, D.C.: A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features. Magn. Reson. Med. 60, 439–448 (2008)
4. Assaf, Y., Basser, P.: Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. NeuroImage 27, 48–58 (2005)
5. Gelderen, P.V., Despres, D., Zijl, P.C.M.V., Moonen, C.T.W.: Evaluation of restricted diffusion in cylinders phosphocreatine in rabbit leg muscle. J. Magn. Reson. Series B 103, 255–260 (1994)
6. Neuman, C.H.: Spin echo of spins diffusing in a bounded medium. J. Chem. Phys. 60, 4508–4511 (1974)
7. Stanisz, G.J., Szafer, A., Wright, G.A., Henkelman, R.M.: An analytical model of restricted diffusion in bovine optic nerve. Magn. Reson. Med. 37, 103–111 (1997)
8. Cook, P.A., Bai, Y., Nedjati-Gilani, S., Seunarine, K.K., Hall, M.G., Parker, G.J., Alexander, D.C.: Camino: Open source diffusion- MRI reconstruction and processing. In: Proc. 14th Meeting of ISMRM, Seattle, WA, USA, vol. 2759 (2006)
9. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. 6, 461–464 (1978)

Multivariate Tensor-Based Brain Anatomical Surface Morphometry via Holomorphic One-Forms

Yalin Wang[1,2], Tony F. Chan[2], Arthur W. Toga[1], and Paul M. Thompson[1]

[1] Lab. of Neuro Imaging, UCLA School of Medicine, Los Angeles, CA 90095, USA
[2] Mathematics Department, UCLA, Los Angeles, CA 90095, USA
ylwang@loni.ucla.edu

Abstract. Here we introduce multivariate tensor-based surface morphometry using holomorphic one-forms to study brain anatomy. We computed new statistics from the Riemannian metric tensors that retain the full information in the deformation tensor fields. We introduce two different holomorphic one-forms that induce different surface conformal parameterizations. We applied this framework to 3D MRI data to analyze hippocampal surface morphometry in Alzheimer's Disease (AD; 26 subjects), lateral ventricular surface morphometry in HIV/AIDS (19 subjects) and cortical surface morphometry in Williams Syndrome (WS; 80 subjects). Experimental results demonstrated that our method powerfully detected brain surface abnormalities. Multivariate statistics on the local tensors outperformed other TBM methods including analysis of the Jacobian determinant, the largest eigenvalue, or the pair of eigenvalues, of the surface Jacobian matrix.

1 Introduction

Surface-based methods have been extensively used to study structural features of the brain, such as cortical gray matter thickness, complexity, and deformation over time [1]. Also, deformation-based morphometry (DBM) [2] directly uses 2D or 3D deformations obtained from the nonlinear registration of brain images to infer local differences in brain volume or shape. Tensor-based morphometry (TBM) [3] tends to examine spatial derivatives of the deformation maps registering brains to a common template, constructing morphological tensor maps such as the Jacobian determinant, torsion or vorticity. DBM, by contrast, tends to analyze 3D displacement vector fields encoding relative positional differences across subjects. One advantage of TBM for surface morphometry is that surfaces are commonly parameterized using grids from which local deformation tensors can be naturally derived - TBM can even make use of the Riemannian surface metric to characterize local anatomical changes.

In computational differential geometry, a holomorphic one-form [4] can be represented as a pair of scalars on each edge of a discrete mesh structure. The holomorphic one-form is an intrinsic, coordinate-free formulation. It provides a practical way to induce conformal parameterizations on surfaces and compute

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 337–344, 2009.

surface-to-surface registrations. The holomorphic one-forms are computed by solving linear systems so the computation is very stable.

In this paper, we present a multivariate TBM framework and apply it to detect abnormal areas on anatomical structures in the brain represented as surfaces, parameterized using differential forms (holomorphic one-forms). We performed three empirical studies of brain abnormalities in Williams syndrome (WS), Alzheimer's Disease (AD), and HIV/AIDS. We studied hippocampal surface deformation associated with AD, and lateral ventricular surface deformation associated with HIV/AIDS. The proposed multivariate TBM detected areas of statistically significant deformation even in relatively small test datasets - one compares 12 subjects with AD to a group of 14 matched healthy controls and the other compares 11 subjects with HIV/AIDS to 8 matched healthy controls. We also detected regions with statistically significant abnormal surface morphology in cortical data from 40 individuals with WS versus 40 matched healthy controls. For comparison, we also applied another three map-based surface statistics to the same three brain anatomical surface datasets. Our goal was to show that the proposed multivariate TBM had more detection power by detecting consistent but more statistically significant areas of abnormal brain structure. Also note that the proposed multivariate TBM framework is simple and general. The Jacobian matrix can be easily computed by Equation 1. Potentially it can take results from any surface registration methods for further morphometry study.

2 Methods

Holomorphic one-forms, a structure used in differential geometry, can be used to generate both canonical conformal parametrization [5] and slit conformal parameterization [6] on 3D anatomical surfaces. The obtained parameterization maximizes the uniformity of the induced grid over the entire domain (see [5,6] for a more detailed algorithm description).

Suppose $\phi : S_1 \rightarrow S_2$ is a map from surface S_1 to surface S_2. The *derivative map* of ϕ is the linear map between the tangent spaces, $d\phi : TM(p) \rightarrow TM(\phi(p))$. In practice, smooth surfaces are usually approximated by triangle meshes. The derivative map $d\phi$ is approximated by the linear map from one face $[v_1, v_2, v_3]$ to another one $[w_1, w_2, w_3]$. First, we isometrically embed the triangle $[v_1, v_2, v_3]$, $[w_1, w_2, w_3]$ onto the plane $\mathbb{R}^2$; the planar coordinates of the vertices of v_i, w_j are denoted using the same symbols v_i, w_j. Then we explicitly compute the linear matrix for the derivative map,$d\phi$, which is the Jacobian matrix of ϕ,

$$d\phi = [w_3 - w_1, w_2 - w_1][v_3 - v_1, v_2 - v_1]^{-1}. \tag{1}$$

In our work, we use multivariate statistics on deformation tensors [7] and adapt the concept to surface tensors. Let J be the derivative map and define the deformation tensors as$S = (J^T J)^{1/2}$. Instead of analyzing shape change based on the eigenvalues of the deformation tensor, we consider a new family of metrics, the "Log-Euclidean metrics" [8]. These metrics make computations on tensors easier to perform, as they are chosen such that the transformed values form a

vector space, and statistical parameters can then be computed easily using the standard formulae for Euclidean spaces.

We apply Hotelling's T^2 test on sets of values in the log-Euclidean space of the deformation tensors. Given two groups of n-dimensional vectors S_i, $i = 1, ..., p$, T_j, $j = 1, ..., q$, we use the Mahalanobis distance M to measure the group mean difference, $M = (log\bar{S} - log\bar{T})\Sigma^{-1}(log\bar{S} - log\bar{T})$, where $\bar{S}$ and $\bar{T}$ are the means of the two groups and Σ is the combined covariance matrix of the two groups.

3 Experimental Results

We applied the multivariate TBM method to various anatomical surfaces extracted from 3D MRI scans of the brain. For registering anatomical surfaces across subjects, we found that conformal slit mapping works well for cortical surface registration because the overall shape of a cortex is close to a sphere and the landmark curve locations are generally similar to each other. On the other hand, holomorphic flow segmentation works better for parameterizing long, cylinder-like shapes, such as hippocampal and lateral ventricular surfaces. In the light of this observation, we used a canonical holomorphic one-form to conformally map hippocampal and lateral ventricular surfaces to a set of planar rectangles (subsection 3.1 and 3.2); and we used slit map conformal parameterization to conformally map cortical surfaces to multiply connected domains (subsection 3.3). Through the parameter domain, we can register surfaces by using a constrained harmonic map [5].

In this paper, the segmentations are regarded as given, and results are from automated and manual segmentations detailed in other prior works [9,10,11].

3.1 Multivariate Tensor-Based Morphometry on Hippocampal Surfaces: Application to Alzheimer's Disease

The hippocampal surface is a structure in the medial temporal lobe of the brain. Parametric shape models of the hippocampus are commonly developed for tracking shape differences or longitudinal atrophy in disease. Many prior studies, e.g., [9], have shown that there is atrophy as the disease progresses. In our method, we leave two holes on the front and back of the hippocampal surface, representing its anterior junction with the amygdala, and its posterior limit as it turns into the white matter of the fornix. It can then be logically represented as an open boundary genus-one surface, i.e., a cylinder. Its canonical holomorphic one-form can be easily computed. By integrating this holomorphic one-form, it can be conformally mapped to a rectangle and registered by using a constrained harmonic map.

Figure 1 (a)-(d) illustrate our experimental results on a group of hippocampal surface models extracted from 3D brain MRI scans of 12 AD individuals and 14 control subjects [9]. After surface registration, we ran a permutation test with 5000 random assignments of subjects to groups to estimate the statistical significance of the areas with group differences in surface morphometry. We also

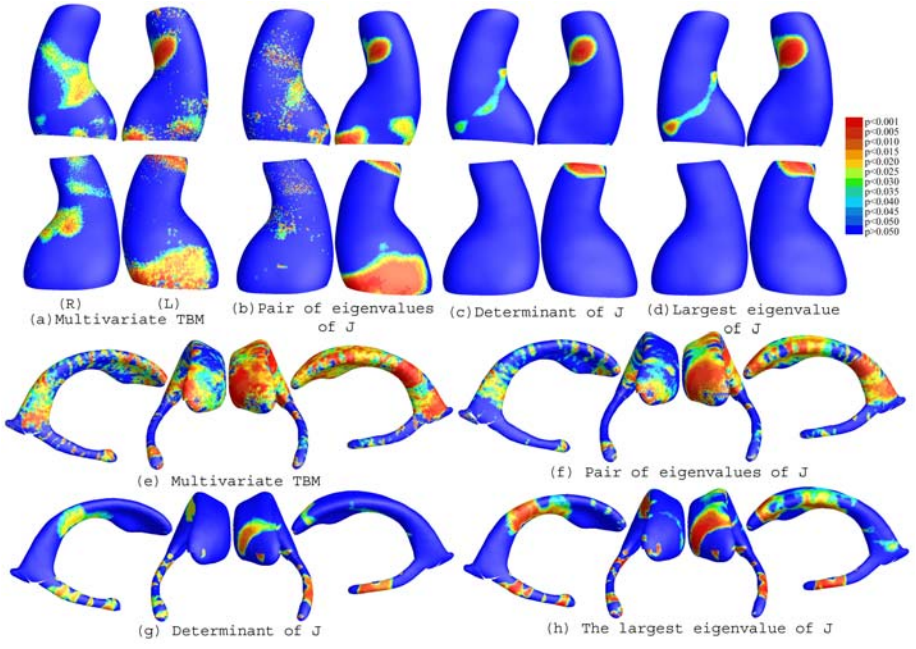


Fig. 1. (a) & (b) illustrate canonical conformal parameterization results. (c)-(j) show statistical p-map results of various TBMs on (1) a group of hippocampal surfaces from 12 AD patients and 14 matched controls((c)-(f)); (2) a group of lateral ventricular surfaces from 11 HIV/AIDS patients and 8 matched controls((g)-(j)). On the color-coded scale, non-blue colors denote the vertices where there is a significant statistical difference, at the $p = 0.05$ level. Multivariate TBM detected anatomical differences more powerfully than other TBM statistics. Overall statistical significance values (corrected for multiple comparisons) are listed in Table 1.

used a statistical threshold of $p = 0.05$ at each surface point to compute the supra-threshold surface area, and we estimate the overall significance of the experimental results by using a non-parametric permutation test to establish an empirical null distribution for this surface area [10]. Although the samples sizes are small, we still detected relatively large statistically significant areas, consistent with prior findings [9]. The overall statistical significance p-values, based on permutation testing (and therefore corrected for multiple comparisons), were 0.0198 for the left hippocampal surface and 0.0410 for the right hippocampal surface (Figure 1 (a)).

3.2 Multivariate Tensor-Based Morphometry of the Ventricular Surface in HIV/AIDS

The lateral ventricles - fluid-filled structures deep in the brain - are often enlarged in disease and can provide sensitive measures of disease progression [11]. Ventricular changes reflect atrophy in surrounding structures, so ventricular measures

and surface-based maps provide sensitive assessments of tissue reduction that correlate with cognitive deterioration in illnesses. However, the concave shape, complex branching topology and extreme narrowness of the inferior and posterior horns have made it difficult for surface parametrization approaches to impose a grid on the entire structure without introducing significant area distortion. To model the lateral ventricular surface, we automatically locate and introduce three cuts on each ventricle. The cuts are motivated by examining the topology of the lateral ventricles, in which several horns are joined together at the ventricular "atrium" or "trigone". We call this topological modeling step, interpreting the ventricles as a set of connected, simpler surfaces, a *topology optimization* operation. The topology optimization helps to enable a uniform parametrization in some areas that otherwise are very difficult to capture with usual parametrization methods. After the topology is modeled in this way, a lateral ventricular surface, in each hemisphere, becomes an open boundary surface with 3 boundaries. We computed the canonical holomorphic one-form [5]. With holomorphic flow segmentation [5], each lateral ventricular surface can be divided into 3 pieces. Although surface geometry is widely variable across subjects, the zero point locations are intrinsically determined by the surface conformal structures, and the partitioning of the surface into component meshes is highly consistent across subjects. The automatic surface segmentation result for a lateral ventricular surface is similar to the manual surface segmentation results used in prior research [11]; even so it improves on past work as it avoids arbitrarily chopping the surface into 3 parts using a fixed coronal plane. After the surface segmentation, each lateral ventricular surface is divided to three surfaces, each topologically equivalent to a cylinder. For each piece, we again applied the holomorphic flow algorithm to it and conformally mapped it to a rectangle. Then we registered each part by a constrained harmonic map. Since all ventricle surfaces are similar and the critical graph is intrinsic to surface, the surface segmentation results are very consistent. It provides a stable surface registration scheme for lateral ventricular surfaces.

In our experiments, we compared ventricular surface models extracted from 3D brain MRI scans of 11 HIV/AIDS individuals and 8 control subjects [11]. After surface registration, we computed the surface Jacobian matrix and applied multivariate tensor-based statistics to study differences in ventricular surface morphometry. We ran a permutation test with 5000 random assignments of subjects to groups to estimate the statistical significance of the areas with group differences in surface morphometry. We also used a statistical threshold of $p = 0.05$ at each surface point to estimate the overall significance of the experimental results by non-parametric permutation test [10]. The experimental results are shown in Figure 1(e). Although sample sizes are small, we still detected large statistically significant areas, consistent with prior findings [11]. The overall statistical significance p-values, based on permutation testing, were 0.0022 for the left lateral ventricle and 0.008 for the right lateral ventricle.

3.3 Multivariate Tensor-Based Morphometry on Cortical Surfaces of Subjects with Williams Syndrome

We also applied our framework to a cortical surface morphometry study of the brain. We analyzed cortical surface models extracted from 3D brain MRI scans of 40 WS individuals and 40 healthy control subjects [10]. We selected a set of 10 landmark curves: the Central Sulcus, Superior Temporal Sulcus Main Body, Inferior Frontal Sulcus, Middle Frontal Sulcus, Inferior Temporal Sulcus, Secondary Intermediate Sulcus, Transverse Occipital Sulcus, Inferior Callosal Outline Segment, Superior Rostral Sulcus, and Subparietal Sulcus. The definitions of these anatomical lines are reported in [12]. After we cut the cortical surface open along the selected landmark curves, a cortical surface became topologically equivalent to an open boundary genus-9 surface. With holomorphic one-forms, the surface can be conformally mapped to an annulus with 8 concentric arcs [6]. Based on surface conformal parameterization, we use the landmark curves as the boundary condition and perform a constrained harmonic map to register the cortical surfaces. For each point on the cortical surface, we ran a permutation test (non-parametric t test) with $5,000$ random assignments of subjects to groups to estimate the statistical significance of the areas with group differences in surface morphometry. Also, given a statistical threshold of p=0.05 at each surface point, we applied permutation test to the overall rejection areas (i.e., using the suprathreshold area statistic) to evaluate the overall significance of the experimental results [9].

After fixing the template parametrization, we used Log-Euclidean metrics to establish a metric on the surface deformation tensors at each point, and conducted a permutation test on the suprathreshold area of the resulting Hotellings T^2 statistics. The statistical map is shown in Figure 2(a). The threshold for significance at each surface point was chosen to be p=0.05. The permutation-based overall significance p values, corrected for multiple comparisons, were p=0.0001 for the right hemisphere and 0.0002 for the left hemisphere, respectively.

3.4 Comparison with Other TBM Methods

To explore whether our multivariate statistics provided extra power when running TBM on the surface data, in each experiment, we also conducted three additional statistical tests based on different tensor-based statistics derived from the Jacobian matrix. The other statistics we studied were: (1) the pair of eigenvalues of the Jacobian matrix, treated as a 2-dimensional vector; (2) the determinant of Jacobian matrix; and (3) the largest eigenvalue of Jacobian matrix. For statistics (2) and (3), we applied a Students t test to compute the group mean difference at each surface point. In case (1), we used Hotelling's T^2 statistics to compute the group mean difference. For the three new statistics, their calculated statistical maps are shown in Figure 1(b)-(d), 1(f)-(h), and 2 (b)-(c), respectively. For each statistic, we also computed the overall p-values (see Table 1). In each experiment, the overall localization and spatial pattern of surface abnormalities detected by different tensor-based surface statistics were highly consistent.

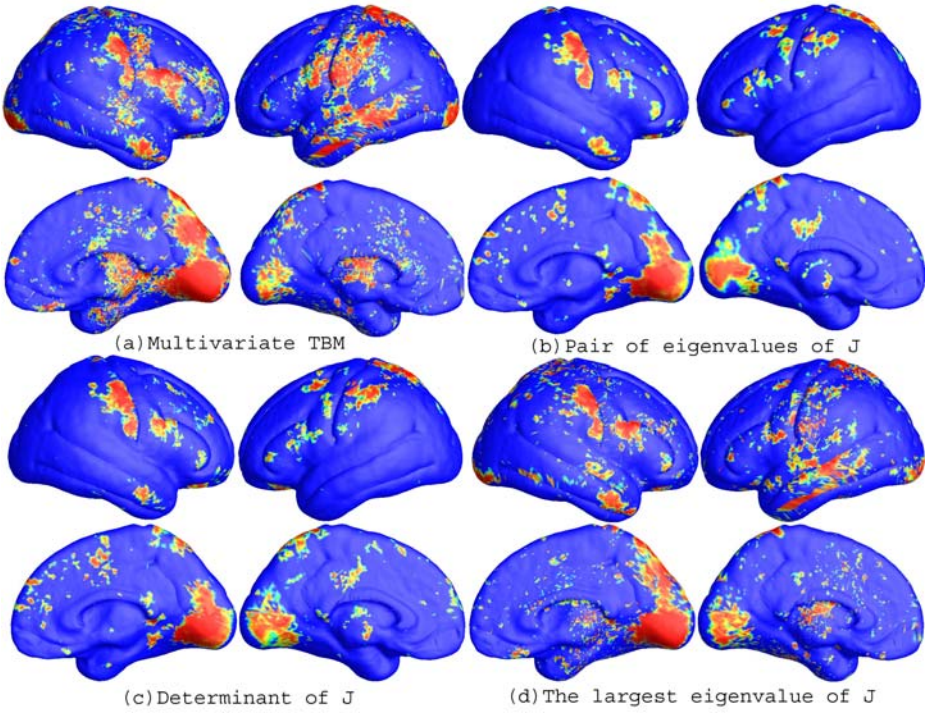


Fig. 2. (a)-(d) illustrate the slit map conformal parameterization on a right hemisphere cortical surface with 10 selected landmark curves. (e)-(h) are illustrative the statistical p-map results of multivariate TBM and other surface TBM for a cortical surface dataset from 40 WS patients and 40 matched control subjects. The color-coded scale is the same as the one in Figure 1, where non-blue colors denote the vertices where there is a significant statistical difference, at the $p = 0.05$ level. Multivariate statistics on the surface Jacobian matrix tend to detect group differences with the greatest effect sizes. Overall statistical significance values (corrected for multiple comparisons) are listed in Table 1.

Table 1. Permutation-based overall significance p value for three experiments. (J is the Jacobian matrix and EV stands for Eigenvalue. To detect group differences, it was advantageous to use the full tensor, or its two eigenvalues together; with simpler local measures based on surface area, group differences were missed.).

	Full Matrix	Determinant of J	Largest EV of J	Pair of EV of J
Left Hippo Surface	**0.0198**	0.1446	0.1016	0.0474
Right Hippo Surface	**0.0410**	0.3600	0.3492	0.0688
Left Vent Surface	**0.0028**	0.0330	0.0098	0.0084
Right Vent Surface	**0.0066**	0.0448	0.0120	0.0226
Left Cortex	**0.0002**	0.1933	0.1627	0.0003
Right Cortex	**0.0001**	0.1366	0.1201	0.0002

The experiments also strongly suggested that the newly proposed multivariate TBM method has more detection power in terms of effect size (and the area with suprathreshold statistics), probably because it captures more directional and rotational information when measuring geometric differences.

Acknowledgments. This work was funded by National Institute of Health through the NIH Roadmap for Medical Research, Grant U54 RR021813 entitled Center for Computational Biology (CCB).

References

1. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis I: segmentation and surface reconstruction. Neuroimage 9, 179–194 (1999)
2. Ashburner, J., Hutton, C., Frackowiak, R., Johnsrude, I., Price, C., Friston, K.: Identifying global anatomical differences: deformation-based morphometry. Human Brain Mapping 6(5-6), 348–357 (1998)
3. Chung, M., Dalton, K., Davidson, R.: Tensor-based cortical surface morphometry via weighted spherical harmonic representation. IEEE Trans. Med. Imag. 27(8), 1143–1151 (2008)
4. Guggenheimer, H.W.: Differential Geometry. Dover Publications (1977)
5. Wang, Y., Gu, X., Hayashi, K.M., Chan, T.F., Thompson, P.M., Yau, S.-T.: Brain surface parameterization using riemann surface structure. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 657–665. Springer, Heidelberg (2005)
6. Wang, Y., Gu, X., Chan, T.F., Thompson, P.M., Yau, S.T.: Conformal slit mapping and its applications to brain surface parameterization. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 585–593. Springer, Heidelberg (2008)
7. Leporé, N., Brun, C., Chou, Y.Y., Chiang, M.C., Dutton, R.A., Hayashi, K.M., Luders, E., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T., Thompson, P.M.: Generalized tensor-based morphometry of HIV/AIDS using multivariate statistics on deformation tensors. IEEE Trans. Med. Imag. 27(1), 129–141 (2008)
8. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Magn. Reson. Med. 56, 411–421 (2006)
9. Thompson, P.M., Hayashi, K.M., de Zubicaray, G.I., Janke, A.L., Rose, S.E., Semple, J., Hong, M.S., Herman, D.H., Gravano, D., Doddrell, D.M., Toga, A.W.: Mapping hippocampal and ventricular change in Alzheimer's disease. NeuroImage 22(4), 1754–1766 (2004)
10. Thompson, P.M., Lee, A.D., Dutton, R.A., Geaga, J.A., Hayashi, K.M., Eckert, M.A., Bellugi, U., Galaburda, A.M., Korenberg, J.R., Mills, D.L., Toga, A.W., Reiss, A.L.: Abnormal cortical complexity and thickness profiles mapped in Williams syndrome. J. Neuroscience 25(16), 4146–4158 (2005)
11. Thompson, P.M., Dutton, R.A., Hayashi, K.M., Lu, A., Lee, S.E., Lee, J.Y., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T.: 3D mapping of ventricular and corpus callosum abnormalities in HIV/AIDS. NeuroImage 31(1), 12–23 (2006)
12. Thompson, P.M., Woods, R.P., Mega, M.S., Toga, A.W.: Mathematical/computational challenges in creating population-based brain atlases. Human Brain Mapping 9(2), 81–92 (2000)

Local White Matter Geometry Indices from Diffusion Tensor Gradients

Peter Savadjiev[1,2], Gordon Kindlmann[3], Sylvain Bouix[1], Martha E. Shenton[1], and Carl-Fredrik Westin[2]

[1] Department of Psychiatry and [2] Department of Radiology
Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
[3] Department of Computer Science, University of Chicago, Chicago, IL, USA

Abstract. We introduce a framework for computing geometrical properties of white matter fibres directly from diffusion tensor fields. The key idea is to isolate the portion of the *gradient* of the tensor field corresponding to local variation in tensor orientation, and to project it onto a coordinate frame of tensor eigenvectors. The resulting eigenframe-centered representation makes it possible to define scalar geometrical measures that describe the underlying white matter fibres, directly from the diffusion tensor field and its gradient, without requiring prior tractography. We define two new scalar measures of (1) fibre dispersion and (2) fibre curving, and we demonstrate them on synthetic and *in-vivo* datasets. Finally, we illustrate their applicability in a group study on schizophrenia.

1 Introduction

Despite the advent of high angular resolution diffusion imaging techniques, diffusion tensor (DT) data continues to be commonly acquired and utilized in a variety of studies in neuroscience and medicine, in particular in clinical settings. Empirically established connections between biological tissue properties and diffusion tensor measures exist. Such tensor measures include e.g. fractional anisotropy (FA), orientation etc., and have been the topic of several studies (e.g. [1,2]). However, few methods consider the differential structure of tensor properties, as introduced in e.g. [3]. In this article we focus on tensor orientation, and perform a differential analysis of diffusion tensor fields which leads to novel methods for the recovery of a variety of white matter geometrical measures, or indices. By isolating the portion of the gradient of the tensor field corresponding to local variation in tensor orientation, and by projecting it onto a coordinate frame of tensor eigenvectors, we achieve an eigenframe-centered representation of local tensor field configurations. Given that the principal eigenvector generally represents the dominant orientation of the underlying fibre population [4], this allows us to define measures of fibre curving and fibre dispersion. The advantage of this approach is the recovery of fibre geometry measures directly from the tensor field, without requiring prior tractography as in e.g. [5].

As scalar measures, these indices can be used to study local fibre organisation in the context of population studies. As a proof-of-concept, we carry out a

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 345–352, 2009.

group study which finds certain differences in local fibre geometry between patients with schizophrenia and normal controls. Other potential applications, to be pursued in future work, include their use in tractography (e.g. as geometrical priors), as well as in a context of registration as geometrical landmarks.

2 Mathematical Model

2.1 Rotation Tangents: Background

Our method is based on the mathematical framework developed in [6]. Let $\boldsymbol{F} : \mathbb{R}^3 \mapsto \mathrm{Sym}_3$ be a diffusion tensor field, such that $\boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{D}$. Here Sym_3 denotes the set of symmetric tensors ($\boldsymbol{D} = \boldsymbol{D}^\top$) in $\mathbb{R}^3 \otimes \mathbb{R}^3$. Consider the tensor rotation function ψ which rotates $\boldsymbol{D}$ with rotation matrix $\boldsymbol{R} \in \mathrm{SO}_3$, the group of rotations on $\mathbb{R}^3$:

$$\psi(\boldsymbol{R}, \boldsymbol{D}) = \boldsymbol{R}\boldsymbol{D}\boldsymbol{R}^\top \ . \tag{1}$$

In [6], the *rotation tangent* $\boldsymbol{\Phi}_i(\boldsymbol{D})$ associated with eigenvector $\boldsymbol{e}_i$ of diffusion tensor $\boldsymbol{D}$ is defined as the change of tensor value due to infinitesimal rotations around $\boldsymbol{e}_i$:

$$\boldsymbol{\Phi}_i(\boldsymbol{D}) = \frac{\partial \psi(\boldsymbol{R}_{\boldsymbol{e}_i}(\phi), \boldsymbol{D})}{\partial \phi}\Big|_{\phi=0} \ , \tag{2}$$

where $\boldsymbol{R}_{\boldsymbol{e}_i}(\phi)$ denotes rotation by angle ϕ around $\boldsymbol{e}_i$. The rotation tangent $\boldsymbol{\Phi}_i(\boldsymbol{D})$ is a second order gradient tensor onto which the tensor field gradient can be projected, in order to obtain three spatial gradients of orientation [6]:

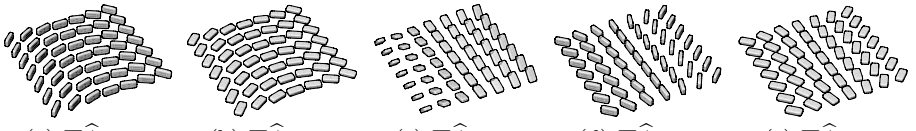
$$\nabla\widehat{\phi}_i(\boldsymbol{x}) = \widehat{\boldsymbol{\Phi}}_i(\boldsymbol{F}(\boldsymbol{x})) : \nabla \boldsymbol{F}(\boldsymbol{x}) \ . \tag{3}$$

Here ":" is the tensor contraction operator (analogous to the vector dot product), $\widehat{\boldsymbol{\Phi}}_i$ are unit-norm rotation tangents, and $\nabla\widehat{\phi}_i$ are vectors that indicate in $\mathbb{R}^3$ the direction in which the tensor orientation around eigenvector $\boldsymbol{e}_i$ varies the fastest. In our work, we compute tensor field gradients as described in [6], by convolving the tensor field with partial derivative kernels of a uniform cubic B-spline.

2.2 Scalar Geometric Measures

The work presented in [7] argues for the representation of white matter fibre geometry (and that of sets of 3D curves in general) in terms of local coordinate frames. The idea is to capture the differential geometry of 3D curves by measuring *changes* in the tangent vector orientation in three mutually orthogonal directions provided by the tangent, normal and bi-normal vectors of a local coordinate frame. The projection of the change of tangent vector orientation in these three directions results in three curvature functions which characterize locally the differential geometry of 3D curve sets.

Motivated by this approach, in this paper we consider the projection of the three $\nabla\widehat{\phi}_i$ (3) into the local coordinate frame provided by the tensor eigenvectors. One can form a total of nine such projections $\nabla\widehat{\phi}_i \cdot \boldsymbol{e}_j$, $i, j \in \{1, 2, 3\}$. We



(a) $\nabla\widehat{\phi}_2 \cdot \boldsymbol{e}_1$ (b) $\nabla\widehat{\phi}_3 \cdot \boldsymbol{e}_1$ (c) $\nabla\widehat{\phi}_2 \cdot \boldsymbol{e}_2$ (d) $\nabla\widehat{\phi}_2 \cdot \boldsymbol{e}_3$ (e) $\nabla\widehat{\phi}_3 \cdot \boldsymbol{e}_2$

Fig. 1. Examples of local tensor field configurations that are characterized by a particular projection of the form $\nabla\widehat{\phi}_i \cdot \boldsymbol{e}_j$

choose the eigenframe as a projection basis due to the biological significance of tensor eigenvectors. It is commonly accepted that the principal eigenvector is aligned with the underlying white matter fibre tract in voxels where the diffusion anisotropy is strong enough (e.g. [4]). We can thus relate patterns of tensor organisation to the geometry of white matter fibres.

Given the assumption that $\boldsymbol{e}_1$ represents the local fibre tangent direction, one can establish correspondences between observed tensor configurations and local fibre configurations. For example, $\nabla\widehat{\phi}_2 \cdot \boldsymbol{e}_1$ and $\nabla\widehat{\phi}_3 \cdot \boldsymbol{e}_1$ measure tangential change in tensor orientation (i.e. change in the tensor field seen by an observer displacing locally in the direction parallel to $\boldsymbol{e}_1$), which is analogous to the classical Frenet curvature, or the tangential curvature [7], of the underlying fibre. On the other hand, the projection of $\nabla\widehat{\phi}_2$ and $\nabla\widehat{\phi}_3$ in the plane spanned by $\boldsymbol{e}_2$ and $\boldsymbol{e}_3$ measures fibre orientation changes in directions orthogonal to $\boldsymbol{e}_1$. This is analogous to the normal and bi-normal curvatures of [7]. Finally, tensor rotations around $\boldsymbol{e}_1$, captured by $\nabla\widehat{\phi}_1$, would correspond to fibre twist.

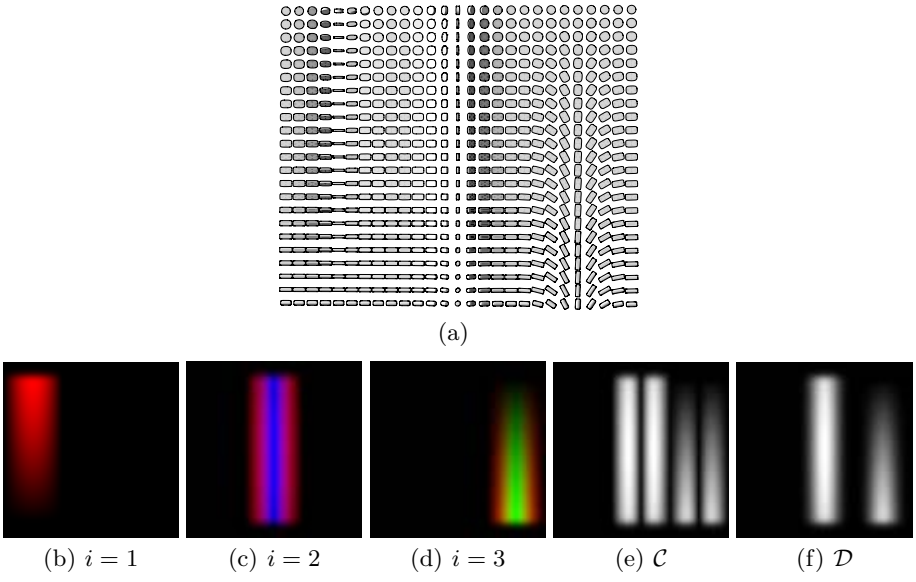
Figure 1 illustrates a variety of local tensor configurations characterized by orientation change, consisting in tensor rotation around an eigenvector $\mathbf{e}_i$ observed in the direction of another eigenvector $\mathbf{e}_j$. For each example configuration, the projection $\nabla\widehat{\phi}_i \cdot \mathbf{e}_j$ will result in a high value (the specific values for i and j are indicated for each case).

Figure 2 illustrates these projections on a 2D synthetic diffusion tensor field with gradients in tensor orientation. Note that the projections are dependent on tensor shape. For instance, rotations around $\boldsymbol{e}_1$ do not result in significant change with nearly cylindrical tensors (i.e. $\lambda_2 = \lambda_3$), hence $\nabla\widehat{\phi}_1 \cdot \boldsymbol{e}_1$ gives smaller values in the bottom left part of Fig. 2(b) as the tensors become more cylindrical. Similarly, as the tensors approach a disk shape (i.e. $\lambda_2 = \lambda_3$), rotation around $\boldsymbol{e}_3$ loses any effect, and $\nabla\widehat{\phi}_3$ and its eigenframe projections go to zero, as can be seen in the upper right part of Fig. 2(d).

We formalize the above observations by defining a local scalar index of fibre curving, $\mathcal{C}$, and a local scalar index of fibre dispersion, $\mathcal{D}$:

$$\textit{Fibre curving:}\quad \mathcal{C}(\boldsymbol{D},\boldsymbol{x}) = \sqrt{\left(\nabla\widehat{\phi}_2\nabla\widehat{\phi}_2^{\top} + \nabla\widehat{\phi}_3\nabla\widehat{\phi}_3^{\top}\right) : \left(\boldsymbol{e}_1\boldsymbol{e}_1^{\top}\right)} \tag{4}$$

$$\textit{Fibre dispersion:}\ \mathcal{D}(\boldsymbol{D},\boldsymbol{x}) = \sqrt{\left(\nabla\widehat{\phi}_2\nabla\widehat{\phi}_2^{\top} + \nabla\widehat{\phi}_3\nabla\widehat{\phi}_3^{\top}\right) : \left(\boldsymbol{e}_2\boldsymbol{e}_2^{\top} + \boldsymbol{e}_3\boldsymbol{e}_3^{\top}\right)} \tag{5}$$



(a) (b) $i = 1$ (c) $i = 2$ (d) $i = 3$ (e) $\mathcal{C}$ (f) $\mathcal{D}$

Fig. 2. (a) A 2D synthetic diffusion tensor field. (b,c,d) Results of projecting $\nabla\widehat{\phi}_i$ onto the local tensor eigenframe. Red: $\nabla\widehat{\phi}_i \cdot e_1$. Green: $\nabla\widehat{\phi}_i \cdot e_2$. Blue: $\nabla\widehat{\phi}_i \cdot e_3$. (e) Curving index. (f) Dispersion index. In (b)-(f), pixel intensity is proportional to index value.

The indices in (4) and (5) combine the rotation tangents $\nabla\widehat{\phi}_2$ and $\nabla\widehat{\phi}_3$ into a single second order tensor $(\nabla\widehat{\phi}_2\nabla\widehat{\phi}_2^{\top} + \nabla\widehat{\phi}_3\nabla\widehat{\phi}_3^{\top})$, which is then projected onto another second order tensor that represents the fibre tangent direction in the case of the curving measure (4), or the plane orthogonal to that direction in the case of (5). The use of outer products and tensor contractions sidesteps the sign ambiguity inherent in both the eigenvectors and the rotation tangents of [6].

In this paper, we do not consider measures involving $\nabla\widehat{\phi}_1$. Unlike the geometry captured by $\nabla\widehat{\phi}_2$ and $\nabla\widehat{\phi}_3$, tensor rotations around $\boldsymbol{e}_1$ do not reflect a change in the fibre tangent direction. Rather, they capture rotations of the diffusion profile *around* the fibre tangent. Such rotations may be particularly sensitive to noise, and their biological meaning is not immediately clear. We thus leave their study for future work.

3 Experiments and Results

3.1 Synthetic Data Validation

We computed the curving (4) and dispersion (5) indices on the synthetic dataset shown in Fig. 2(a). It is divided in three vertical regions of rotation, one for each eigenvector. Eigenvalue mean and variance are constant throughout the field, and mode [6] increases smoothly from top to bottom. The results are shown in Fig. 2(e,f). As expected, the fanning tensor configurations in the central and

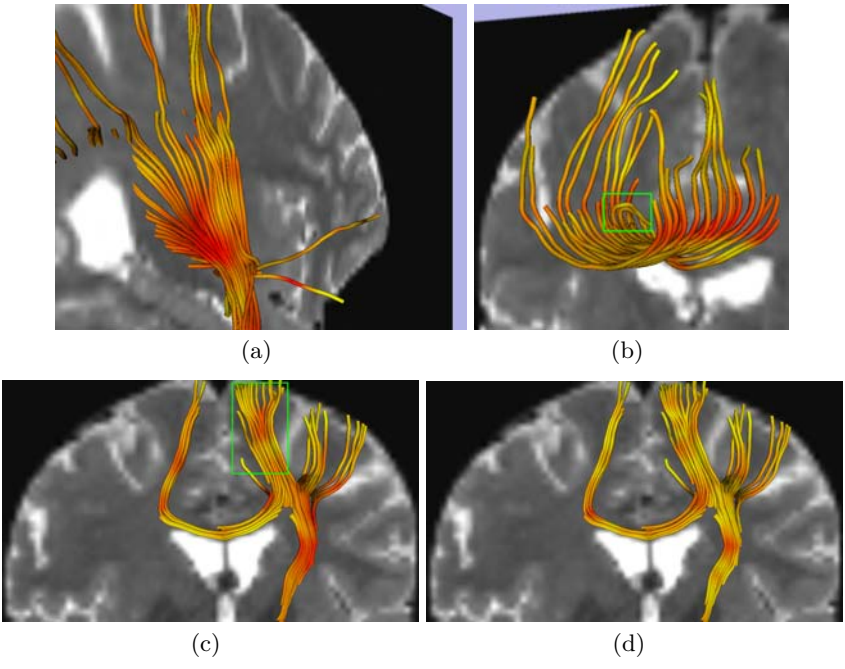


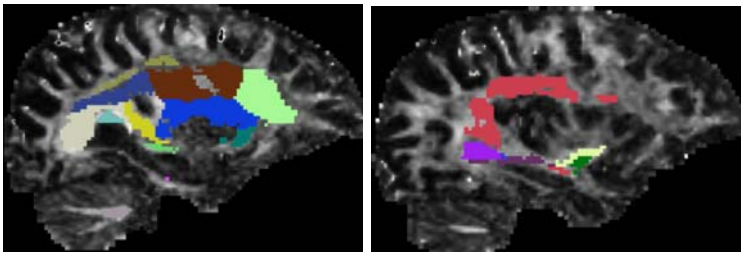
Fig. 3. Streamline DT tractography colored by dispersion index (a,c) and curving index (b,d). Red indicates high index values. The T2-weighted image is shown in the background for reference. Note the consistent behaviour of the pre-computed dispersion and curving indices with respect to the tractography results. The regions in the green boxes are discussed further in the text. Some of the strong dispersion regions in the internal capsule (c) are due to fibre fanning orthogonal to the image plane.

right-hand side vertical bands have a high dispersion index. The immediately adjacent regions exhibit a high curving index, which is consistent with the underlying geometry. Note that the two measures are not mutually exclusive, and some locations can have both high dispersion and high curving indices.

3.2 *In-vivo* Data Validation

Diffusion-weighted images were acquired on a 3T scanner (General Electric Company, Milwaukee, WI, USA) using an echo planar imaging (EPI) sequence, with a double echo option to reduce eddy-current related distortions. To reduce the impact of EPI spatial distortion, an 8 Channel coil and ASSET with a SENSE-factor of 2 were used. The acquisition consisted in 51 directions with b=900, and 8 baseline images with b=0. The scan parameters were: TR=17000 ms, TE=78 ms, FOV=24 cm, 144 $\times$ 144 encoding steps, 1.7 mm slice thickness. A total of 85 axial slices covering the whole brain were acquired.

The geometric indices were computed only in voxels where the FA value was above 0.2, in order to avoid tensors with uncertain eigenvector directions, and



(a) (b)

Fig. 4. Sagittal slices through a patient's right hemisphere, with the ROIs overlaid on an FA map. The results reported in Fig. 5(a,b) pertain to the frontal corona radiata ROI, shown in light green in (a). Those presented in Fig. 5(c,d) pertain to a segmentation of the uncinate, which combines the light and dark green ROIs in (b). Note the ROIs are 3D volumes, only a slice through which is visible here.

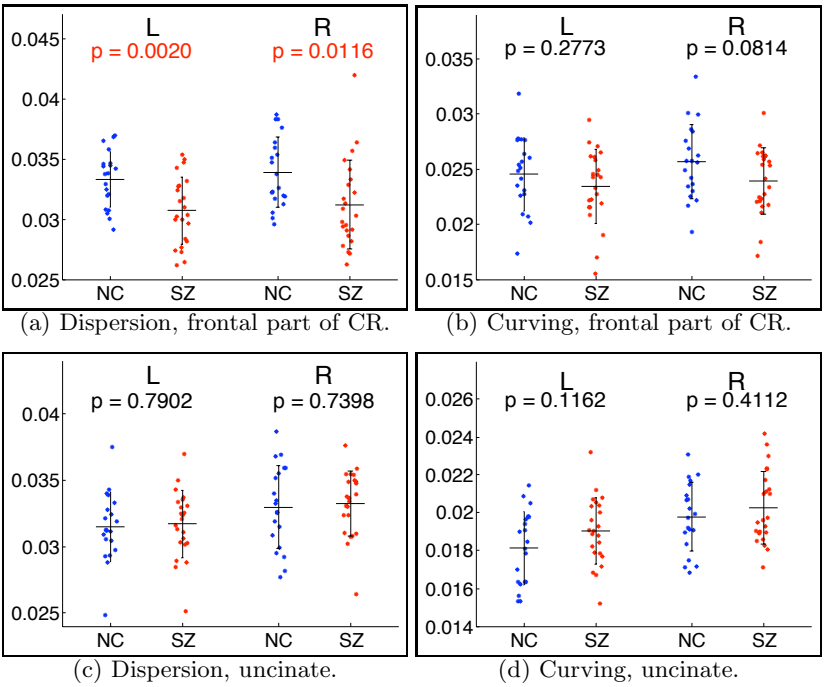
CSF or grey matter voxels. Prior to computation of the geometrical indices, tensors were normalized by dividing by tensor norm (see e.g. [6] for a definition of tensor norm). After computation of the indices, standard streamline tractography [8] was run to help visualize in three dimensions the behaviour of the indices relative to well-known fibres. The resulting tracts were colored with the pre-computed scalar indices, as shown in Fig. 3. Note how the dispersion index captures the local fanning of fibres passing through the internal capsule in Fig. 3 (a). The fibres run parallel in the bottom part of the green box in Fig. 3(c) and the dispersion index is low. In the upper part of the box, however, the fibres fan, which is characterized by higher index values. While assessing visually these results, it is important to remember that fibre dispersion is a 3D phenomenon.

Note how the curving index results in Fig. 3(b,d) are consistent with the geometry of the recovered fibre tracts. An exception is seen in the small region enclosed by a green box in (b), where the low index values do not appear to match the geometry of the strongly curved fibre. It appears here that the tractography algorithm wrongly "jumped" onto the cingulum after initially following a callosal fibre. This example shows how our geometrical indices could potentially be used as priors to direct tractography and remove incorrect paths.

3.3 Group Study on Schizophrenia

Diffusion MRI data from 20 normal controls (NC) and 23 schizophrenic patients (SZ) was acquired and preprocessed as described in Section 3.2. The deep white matter structures were segmented in regions of interest (ROI) by registering the ICBM-DTI-81 atlas [9] with the diffusion baseline of each subject using nonlinear registration [8].

We computed the dispersion (5) and curving (4) indices in ROIs that segment the frontal part of the corona radiata (CR), as well as the uncinate fasciculus (UN), in both hemispheres. A slice through each ROI in one patient is illustrated



(a) Dispersion, frontal part of CR. (b) Curving, frontal part of CR.

(c) Dispersion, uncinate. (d) Curving, uncinate.

Fig. 5. Comparison between NC and SZ through scatter plots of mean index value for each individual. Top row: results from the frontal part of the corona radiata. Bottom row: results from the uncinate fasciculus. Left column: dispersion index. Right column: curving index. Blue: NC. Red: SZ. Horizontal black lines: population mean. The error bars indicate $\pm$ 1 standard deviation. L: left hemisphere. R: right hemisphere. The p-values are computed with a paired-value T-test, and those that satisfy a significance threshold of 0.05 are given in red font.

in Fig. 4. The CR is characterized by dispersion, whereas the UN is a curving tract. We thus hypothesized that the dispersion index in the CR case and the curving index in the UN case may reveal certain population differences. The results are presented in Fig. 5. The difference between the NC and SZ populations is significant in the case of the dispersion values in the frontal CR, shown in Fig. 5(a). The curving index comparison in Fig. 5(b,d) shows a trend towards the SZ group having higher index values in the UN, and lower values in the CR. Thus, the population differences are reflected mainly in the dispersion index in the frontal CR. As for the curving index, it reveals an interesting left vs. right asymmetry in the UN.

These results are preliminary, and further study is required before clinical significance claims can be made. For instance, a better ROI definition scheme may be required. Nevertheless, these results serve as a proof of concept and demonstrate that geometrical indices can recover population trends.

4 Conclusion

This paper introduced novel scalar indices of white matter dispersion and curving. Since they are computed directly from the tensor field, without requiring prior tractography, these indices simplify the geometrical analysis of white matter and make it insensitive to possible tractography errors. Traditionally, clinical DTI studies focus on anisotropy, but our results demonstrate that geometrical measures may also be important. Clearly, our region-based comparison results depend on the quality of white matter segmentation. We will address this issue as part of future clinical comparison studies. In work not reported here, we experimented with *in-vivo* datasets where all tensors with anisotropy higher than a threshold were transformed to a cylindrical shape, by assigning the same eigenvalues to all such tensors, with the idea of removing the influence of tensor shape on the geometrical index values. The results were surprisingly similar to those reported here, and will be discussed in future extensions of this work. Finally, we intend to address applications in fibre tractography and DT image registration, as well as the extension of our indices to a multi-scale framework.

Acknowledgments. Work supported by NIH grants R01MH074794, R01MH50740, K05MH070047, P50MH080272-01, P41RR13218, U54EB005149, U41RR019703, Department of Veteran Affairs Merit Awards, VA Schizophrenia Center.

References

1. Rovaris, M., Filippi, M.: Diffusion tensor MRI in multiple sclerosis. J. Neuroimaging 17(s1), 27s–30s (2007)
2. Kubicki, M., McCarley, R., Westin, C.F., Park, H.J., Maier, S., Kikinis, R., Jolesz, F.A., Shenton, M.E.: A review of diffusion tensor imaging studies in schizophrenia. J. Psychiatric Research 41, 15–30 (2007)
3. Pajevic, S., Aldroubi, A., Basser, P.J.: A continuous tensor field approximation of discrete DT-MRI data for extracting microstructural and architectural features of tissue. J. Magn. Resonance 154, 85–100 (2002)
4. Lin, C.P., Tseng, W.Y.I., Cheng, H.C., Chen, J.H.: Validation of diffusion tensor magnetic resonance axonal fiber imaging with registered manganese-enhanced optic tracts. NeuroImage 14, 1035–1047 (2001)
5. Batchelor, P.G., Calamante, F., Tournier, J.D., Atkinson, D., Hill, D.L.G., Connelly, A.: Quantification of the shape of fiber tracts. Magn. Res. in Medicine 55, 894–903 (2006)
6. Kindlmann, G., Ennis, D.B., Whitaker, R.T., Westin, C.F.: Diffusion tensor analysis with invariant gradients and rotation tangents. IEEE Trans. Medical Imaging 26(11), 1483–1499 (2007)
7. Savadjiev, P., Zucker, S.W., Siddiqi, K.: On the differential geometry of 3D flow patterns: Generalized helicoids and diffusion MRI analysis. In: Proc. IEEE Intl. Conf. on Computer Vision, ICCV 2007 (2007)
8. http://www.slicer.org
9. http://loni.ucla.edu/Atlases/

i-BRUSH: A Gaze-Contingent Virtual Paintbrush for Dense 3D Reconstruction in Robotic Assisted Surgery

Marco Visentini-Scarzanella[1], George P. Mylonas[1,2], Danail Stoyanov[2], and Guang-Zhong Yang[1,2]

[1] Royal Society/Wolfson Foundation MIC Laboratory
[2] Institute of Biomedical Engineering,
Imperial College London, London SW7 2AZ, United Kingdom
{marco.visentini-scarzanella02,george.mylonas,danail.stoyanov,
g.z.yang}@imperial.ac.uk

Abstract. With increasing demand on intra-operative navigation and motion compensation during robotic assisted minimally invasive surgery, real-time 3D deformation recovery remains a central problem. Currently the majority of existing methods rely on salient features, where the inherent paucity of distinctive landmarks implies either a semi-dense reconstruction or the use of strong geometrical constraints. In this study, we propose a gaze-contingent depth reconstruction scheme by integrating human perception with semi-dense stereo and *p-q* based shading information. Depth inference is carried out in real-time through a novel application of Bayesian chains without smoothness priors. The practical value of the scheme is highlighted by detailed validation using a beating heart phantom model with known geometry to verify the performance of gaze-contingent 3D surface reconstruction and deformation recovery.

1 Introduction

In robotic assisted minimally invasive surgery (MIS), the ability to reconstruct 3D tissue deformation in real-time is required for the prescription of dynamic active constraints [1], motion stabilisation, and intra-operative image guidance. Recently, it has been shown that it is possible to recover the 3D shape and morphology of soft-tissue surfaces using a stereo-laparoscope for anatomical regions with distinctive appearance. However, to recover 3D structure from homogeneous tissue areas with limited texture details, strong geometrical constraints such as hierarchical piecewise bilinear maps, B-splines and thin-plate splines have been used to provide dense disparity maps [2-4]. With these approaches, surface discontinuities and specular highlights are significant problems as is the assumption that the scene is composed of a single smooth surface. Alternative techniques based on photometric algorithms forgo the need for explicit feature mapping: by assuming an *a priori* physical image formation model, particularly the co-axial arrangement of the endoscope camera and light-source, dense 3D reconstruction is feasible. The uniqueness of viscosity solutions to a Hamilton-Jacobi equation for the perspective projection case can be established but most methods assume uniform albedo and a Lambertian reflection model [5]. Systems integrating multiple visual cues through inference over Markov Random Fields

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 353–360, 2009.

(MRF) for 3D reconstruction have been demonstrated [6, 7]. However, their computational complexity limits their applicability in scenarios requiring real-time performance and high precision.

Recently, the concept of gaze-contingent perceptual docking through binocular eye tracking has attracted extensive interest [8]. Current results suggest that 3D recovery and motion stabilisation in the fovea is sufficient, while stabilisation in the periphery does not affect user and visual perceptual quality [9]. The strength of the gaze-contingent method is that it relies on the innate ability of the human vision, independently from the geometrical characteristics of the scene presented.

The purpose of this paper is to present an interactive gaze-contingent approach to dynamically reconstruct the 3D structure of the surgeon's area of interest. Surface recovery is formulated as an inference problem combining stereo disparity estimation and surface shading for increased robustness. Computational complexity is limited by considering only the surgeon's area of interest, and by a novel usage of Bayesian chains instead of MRFs for propagation without smoothness constraints. A fast regularisation scheme ensures all information available is used without loopy propagation, enabling real-time operation.

The proposed method, called *i*-BRUSH (Bayesian Reconstruction Using Stereo and sHading), is validated on a da Vinci surgical robot using a beating heart phantom with known geometrical and temporal ground truth data for surface reconstruction and tissue deformation recovery.

2 Method

The proposed method consists of four distinct processes. First, the surgeon's fixation point is localised on the left and right channels of the input video stream from the endoscope through the use of a binocular eye-tracker fully integrated with the da Vinci surgical robot's master console. The eye-tracker consists of a pair of near infrared sensitive cameras, an array of externally switchable miniaturised IR emitting diodes at 940nm and a pair of dichroic beam splitters with cut-off wavelength at 750nm. This allows both eyes to be tracked at 50 fps after a short eye calibration procedure as outlined in the work by Mylonas *et al.* [10].

Then, stereo reconstruction is employed to provide a semi-dense estimation of the surface within the small area of high visual acuity corresponding to roughly 3 to 5 degrees of visual angle. This is achieved by using a Shi-Tomasi detector followed by pyramidal Lucas-Kanade tracker for matching. This sparse 3D information is then locally propagated, yielding a semi-dense reconstruction. Independently from the stereo reconstruction process, local surface orientation expressed as the surface directional gradient vectors (p_i, q_i) is estimated for each pixel m_i by approximating the image irradiance equation based on Lambertian reflection and the geometrical constraints of the endoscope camera and light-sources as described in [11].

The final stage of the proposed method focuses on fusion of the semi-dense stereo map with the *p*-*q* surface gradient to obtain a dense reconstruction of the area. To this end, a Bayesian inference scheme is used. While a grid network would have to be solved through approximate methods, thus introducing excessive computational burden, a polytree is used instead for efficient inference at each image element. This

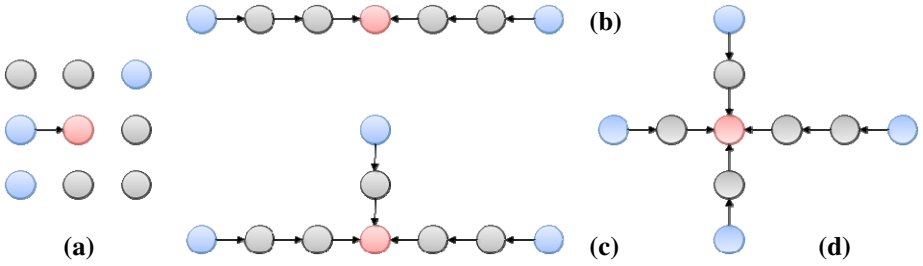


Fig. 1. Network topologies depending on the number of stereo-reconstructed points present along the horizontal and vertical directions. Stereo reconstructed points are shown in blue, the point to be inferred in red and points with only *p-q* information are shown in grey. **(a)** The network contains only one stereo point along the horizontal or vertical direction; **(b)** two points along the same direction; **(c)** three chains; and **(d)** four chains, respectively. The chains may not be symmetrical as the first stereo pixel along a given direction is taken.

structure consists of a maximum of four Bayesian chains in a star configuration, each starting from a separate stereo-reconstructed point and terminating at the pixel to be inferred. These Bayesian chains represent the shortest direct path between the point to be inferred and the stereo-reconstructed points, with the terminating points imposing depth constraints and the nodes along the chain providing the natural encoding of the *p-q* gradient derived from local surface shading. Four different topologies are possible depending on the number of stereo points, and these are illustrated in Fig. 1.

The problem of exact inference along causal polytrees, such as the ones illustrated in Fig. 1, is well studied and requires a computationally intensive Belief Propagation (BP) message passing scheme. It can be shown that without the use of smoothness constraints, the problem can be reduced to a simple recursive scheme.

Let us define the problem of finding the Maximum A Posteriori (MAP) estimate of the distance from the camera centre D_k for the k^{th} element along one chain, with upstream information being propagated from N chains. The information propagated consists of ΔD^{n+} and pq^{n+}, as the depth gradient and *p-q* information propagated from the n^{th} upstream chain. Since a node is conditionally independent on all others given its own Markov blanket, the MAP estimate can be written as:

$$\tilde{D}_k = \arg\max\left\{P'(D_k)\right\} = \arg\max_{D_k \in \mathbb{R}}\left\{P\left(D_k \mid \Delta D_{k-1}^{n+}, pq_k, pq_{k-1}^{n+}\right)\right\} \tag{1}$$

Separating upstream and element evidence, and setting α as a normalisation constant:

$$P'(D_k) = \alpha\left(P(D_k \mid pq_k)\prod_{n^+ \in N} P\left(D_k \mid \Delta D_{k-1}^{n+}, pq_{k-1}^{n+}\right)\right) \tag{2}$$

which can be solved with belief propagation requiring likelihood information to be passed to all nodes along the chain. Since $D_k = D_{k-1} + \Delta D_k$, (2) can be rewritten as:

$$P'(D_k) = \alpha\left(P(\Delta D_k \mid pq_k)\prod_{n^+ \in N} P\left(\Delta D_k \mid \Delta D_{k-1}^{n+}, pq_{k-1}^{n+}\right)\right) + \beta\sum_{n^+ \in N} D_{k-1}^{n+} \tag{3}$$

The normalisation constant β is equal to the reciprocal of the number of chains. As the surface gradient at the k^{th} node is independent from its neighbours' gradient -unless surface smoothness is assumed- the gradient can be inferred from the likelihood relationship with the instantiated value of the surface gradient:

$$P'(D_k) = P(\Delta D_k \mid pq_k) + \beta \sum_{n^+ \in N} D_{k-1}^{n+} \tag{4}$$

The recursion can be expanded up to the terminal points of the chain, which propagate the initial stereo values. The MAP estimate of the gradient can then be obtained from likelihood data correlating surface gradient and *p-q* values combined with depth information from the terminal nodes. The matrices are obtained for each image frame by discretising the range of values, obtained from the stereo and *p-q* measurements, into a number of bins sufficient for maintaining precision while guaranteeing real-time performance.

Local inaccuracies in the *p-q* and stereo reconstructions are mitigated by the use of global per-frame statistics, however, errors in the instantiated values during inference can result in a cumulative error. By assuming that the error at each step is normally distributed, the estimation error of $\tilde{D}_k$ can be reduced by setting it as the weighted average of the estimates D_{k-1}^{n+} from the N converging chains. The weighting factor is determined by the number of nodes K_n along the chain as a fraction of the total number of nodes K from all chains. Thus, for the point to be inferred, the normalisation constant β can be set as a weighted average accordingly, yielding:

$$\tilde{D}_k = \arg\max \left\{ P(\Delta D_k \mid pq_k) + \frac{1}{Z} \sum_{n^+ \in N} \frac{K - K_n}{K} D_{k-1}^{n+} \right\} \tag{5}$$

The normalisation constant *1/Z* ensures the weights add up to one. While (5) can be used to recover the depth of the point m_i to be inferred, it does not address how to use the other estimates obtained along the chains. Such information can be used to update the local frame statistics if it is regularised to reflect the new weighted average estimate of m_i. To regularise the estimates without explicit smoothness constraints, the chain can be modelled as a spring series fixed at one end. The stiffness constant of each spring is inversely proportional to the number of nodes in the chain, since the confidence of the estimate decreases with each inferential step, and the one-dimensional relaxed position of the k^{th} spring joint along the n^{th} chain is equal to its MAP estimate. The difference between the weighted average and the estimate for each chain Δm_i^n can be then represented as the result of a constant force F acting on the spring system. The force is distributed along the springs according to their relative stiffness constants, where the compressed position of the k^{th} spring joint is equal to the new depth estimate at the corresponding node along the chain. The force is calculated from the displacement of a single equivalent spring:

$$F = x_{eq} p_{eq} = \Delta m_i^n \left(\sum_{k=1}^{K_n} \frac{1}{p_k} \right)^{-1} = \Delta m_i^n \left(\sum_{k=1}^{K_n} k \right)^{-1} \tag{6}$$

Since the force is constant at each joint, the displacement of each node along the n^{th} chain Δm_k^n can be calculated by equating the force on the equivalent spring with the force at each joint and solving for Δm_k^n. The likelihood matrices for the elements along each inference chain are then updated to integrate the new evidence for subsequent iterations involving any subset of the inferred nodes. Once all points not previously reconstructed have been inferred, their depth is converted into 3D coordinates by projecting a ray of length equal to the estimated depth from the camera centre through the 2D point m_i on the retinal plane.

3 Experimental Setup and Results

To evaluate the accuracy of the proposed depth reconstruction scheme, a silicone phantom heart (Chamberlain Group, MA, USA) has been recorded at 25 fps using the stereo laparoscope from a da Vinci system integrated with the 3D binocular eye-tracker as described earlier. To facilitate the registration process during validation, the phantom heart was augmented with 15 silicone fiducials visible on its surface and scanned with a Siemens Somatom Sensation 64 CT scanner. The image resolution was 0.41×0.41×0.5 mm and the temporal resolution was 0.33s. In this experiment, the heart rate was set to 90bpm, yielding twenty 3D volumes covering the entire cardiac cycle. To ensure the accuracy of the ground truth data, the cardiac volumes and the fiducials were segmented interactively and the spatial coordinates of the fiducials were tracked. An image of the phantom together with a 3D rendering of one cardiac phase indicating the location of the fiducials is shown in Fig. 2.

Two sets of experiments were performed to illustrate the potential applications of the proposed method. First, the operator was asked to fixate on the surface of the static phantom describing continuous paths in order to demonstrate the method's accuracy in

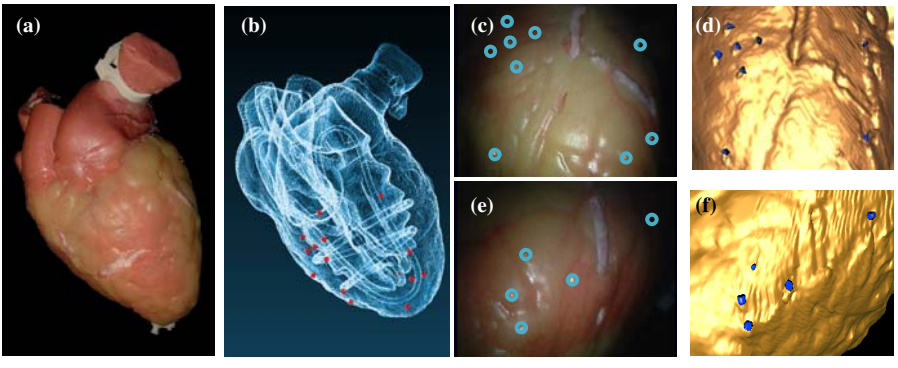


Fig. 2. (a) The heart phantom and **(b)** its corresponding 3D rendering from the 4D CT data. Fiducial locations are highlighted in red on the epicardial surface. **(c)**, **(e)** Image frames from LAD and left ventricle sequences, fiducials highlighted in light blue. **(d)**, **(f)** Registration of ground truth data to camera coordinates of **(c)**, **(e)** respectively using manually-tracked fiducial locations.

recovering 3D information for large smooth featureless areas. Second, the operator was required to perform fixations on different regions of the epicardial surface of the beating heart phantom to evaluate the accuracy of real-time dynamic motion recovery.

For each experiment, two video sequences were recorded from different viewpoints, the first focusing on the left ventricle and the second surrounding the midsegment of the left anterior descending artery (LAD) The global pose alignment of the heart with the CT data for validation was based on the fiducials located on the epicardial surface. To calculate the optimal registration between the two coordinate systems knowing the correspondence between fiducials, the absolute orientation method by Horn [12] was used. The average alignment error was estimated to be of 1.30 mm and 1.18 mm for the sequences **(c)** and **(e)** shown in Fig. 2.

Static Surface Reconstruction. For the static sequences, paths of different lengths were visually delineated on the epicardial surface, both on uniform areas and on parts of the surface presenting distinctive landmarks. To estimate the performance numerically, the reconstruction error for each 3D point $\mathbf{M_i}$ is given as the distance between $\mathbf{M_i}$ and the closest point belonging to the ground truth set lying along the ray passing through $\mathbf{M_i}$ and the camera origin. To this end, a ray-casting method has been implemented to recover the barycentric coordinates of the intersection point between the ray and a triangular facet described by three vertices on the CT volume. In Fig. 3, the paths reconstructed are overlaid on the original image frames together with a graphical representation of the error distribution for each frame.

It is evident that the frequencies in Fig. 3(b) and (e) show that most reconstruction errors lie in the range 0-3.5mm with the outliers present in some frames significantly affecting the standard deviation of the error shown in Figs. 3(c) and 3(f). The outliers occur due to occasional drifts between the recorded left and right fixations producing incorrect stereo matches. However, as a significant proportion of the reconstructed area consists of overlapping regions between consecutive frames, the impact of outliers can be easily removed by averaging of all the estimates obtained for a given point. The average error for the overall reconstructed surfaces in Figs. 3(a) and (d) is of 2.30 mm and 2.03 mm respectively, with a standard deviation of 2.51 and 3.64 mm. Such values are in line with the initial alignment error estimated in the previous section, suggesting the potential of the method for accurate dense 3D estimation of static scenes.

Dynamic Depth Recovery. The proposed method was also applied to motion estimation on the beating phantom. Evaluation was performed as previously indicated with the temporally aligned and spatially registered CT data. The accuracy of the reconstructed area was in line with that of the static experiments. The per pixel error appears to be uniformly distributed between positive and negative values, yielding an estimated average distance closer to the ground truth. Discrepancies between ground truth and recovered motion are a consequence of the inability to visually track fast moving objects with small displacements, and it would be alleviated at lower beat rates. A summary of the overall error assessment is presented in Table 1.

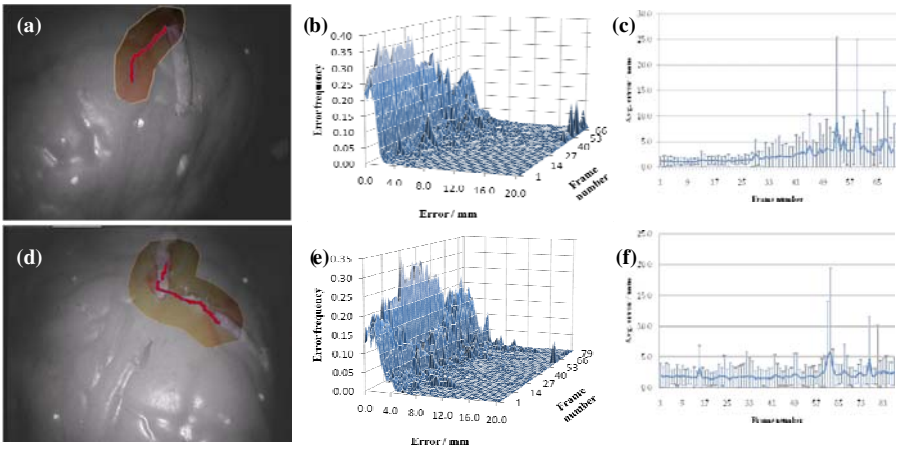


Fig. 3. **(a)** Left ventricle and **(d)** LAD section sequences, the traces of fixations along paths are shown in red. **(b)**, **(e)** Per-frame reconstruction error frequency of **(a)**, **(d)**. **(c)**, **(f)** Average and standard deviation of reconstruction error for each frame from **(a)**, **(d)**.

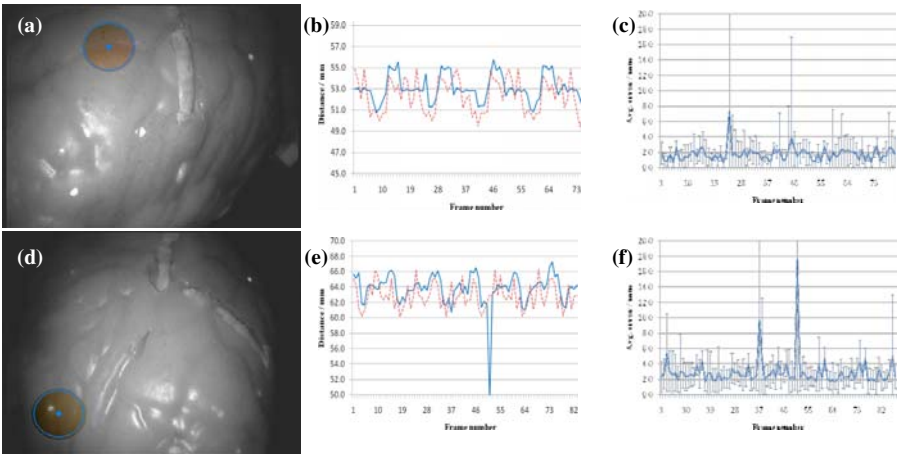


Fig. 4. **(a)** Left ventricle and **(d)** LAD sequences, the fixation area is highlighted in blue. **(b)**, **(e)** Per-frame average depth of the reconstructed and ground truth fixation area. **(c)**, **(f)** Per-frame average error and standard deviation.

Table 1. Summary of results for static and dynamic sequences

Sequence	Length (frames)	Avg. distance (mm)	Avg. error (mm)	Standard deviation	Avg. error (%)
Left ventricle (static)	70	53.493	2.302	2.509	4.30
Left ventricle (dynamic)	80	53.081	1.809	2.036	3.41
LAD (static)	85	67.837	2.029	3.641	2.99
LAD (dynamic)	100	63.793	2.936	2.331	4.60

4 Discussion and Conclusions

In this paper, we have presented a new framework for intra-operative 3D reconstruction from multiple depth cues for addressing the problem of dense reconstruction without imposing prior geometric constraints. The reconstructed surface is limited to the surgeon's main area of interest, thus facilitating real-time implementation with no detrimental effect on perceptual quality. It has been further demonstrated that with the proposed method the removal of explicit smoothness constraints has no adverse effect on the reconstruction accuracy while reducing computational complexity. Experimental results from a phantom model with known ground truth highlight the value of the technique for both static and dynamic scenes, with potential applications in image guidance, motor channelling and motion compensation. Future work will explore the inclusion of multiple reflectance models to better represent the material properties and temporal cues for added robustness.

References

1. Mylonas, G., Kwok, K.W., Darzi, A., Yang, G.-Z.: Gaze-Contingent Motor Channelling and Haptic Constraints for Minimally Invasive Robotic Surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 676–683. Springer, Heidelberg (2008)
2. Lau, W.W., Ramey, N.A., Corso, J.J., Thakor, N.V., Hager, G.D.: Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 494–501. Springer, Heidelberg (2004)
3. Richa, R., Poignet, P., Liu, C.: Efficient 3D Tracking for Motion Compensation in Beating Heart Surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 684–691. Springer, Heidelberg (2008)
4. Stoyanov, D., Darzi, A., Yang, G.-Z.: Dense 3D Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 41–48. Springer, Heidelberg (2004)
5. Prados, E., Faugeras, O.: Shape from shading: a well-posed problem? In: IEEE Computer Society Conference on CVPR, vol. 872 (2005)
6. Yuille, A., Kersten, D.: Vision as Bayesian inference: analysis by synthesis? Trends in Cognitive Sciences 10(7), 301–308 (2006)
7. Lo, B.P.L., Visentini Scarzanella, M., Stoyanov, D., Yang, G.-Z.: Belief Propagation for Depth Cue Fusion in Minimally Invasive Surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 104–112. Springer, Heidelberg (2008)
8. Mylonas, G.P., Darzi, A., Yang, G.-Z.: Gaze Contingent Depth Recovery and Motion Stabilisation for Minimally Invasive Robotic Surgery. In: Yang, G.-Z., Jiang, T.-Z. (eds.) MIAR 2004. LNCS, vol. 3150, pp. 311–319. Springer, Heidelberg (2004)
9. Mylonas, G., Stoyanov, D., Darzi, A., Yang, G.-Z.: Assessment of Perceptual Quality for Gaze-Contingent Motion Stabilization in Robotic Assisted Minimally Invasive Surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 660–667. Springer, Heidelberg (2007)
10. Mylonas, G.P., Darzi, A., Yang, G.-Z.: Gaze-contingent control for minimally invasive robotic surgery. Computer Aided Surgery 11(5), 256–266 (2006)
11. Rashid, H.U., Burger, P.: Differential algorithm for the determination of shape from shading using a point light source. Image Vision Comput 10(2), 119–127 (1992)
12. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. J. Opt. Soc. Am. A 4(4), 629–642 (1987)

Targeting Accuracy under Model-to-Subject Misalignments in Model-Guided Cardiac Surgery

Cristian A. Linte, John Moore, Andrew D. Wiles, Chris Wedlake, and Terry M. Peters

Imaging Research Laboratories, Robarts Research Institute,
Biomedical Engineering Graduate Program, University of Western Ontario,
Department of Medical Biophysics, University of Western Ontario,
London, Ontario, Canada
{clinte,tpeters}@imaging.robarts.ca

Abstract. In image-guided interventions, anatomical models of organs are often generated from pre-operative images and further employed in planning and guiding therapeutic procedures. However, the accuracy of these models, along with their registration to the subject are crucial for successful therapy delivery. These factors are amplified when manipulating soft tissue undergoing large deformations, such as the heart. When used in guiding beating-heart procedures, pre-operative models may not be sufficient for guidance and they are often complemented with real-time, intra-operative cardiac imaging. Here we demonstrate via *in vitro* endocardial "therapy" that ultrasound-enhanced model-guided navigation provides sufficient guidance to preserve a clinically-desired targeting accuracy of under 3 mm independently of the model-to-subject misregistrations. These results emphasize the direct benefit of integrating real-time imaging within intra-operative visualization environments considering that model-to-subject misalignments are often encountered clinically.

1 Introduction

The development of minimally invasive alternatives to conventional cardiac interventions are under active investigation to reduce trauma and recovery time [1,2,3]. To accelerate the progress toward the ultimate, least invasive treatment approach — beating heart intracardiac therapy, robust intra-operative visualization is crucial. These environments need to faithfully represent the surgical field, assist the surgeon with correctly identifying the targets to be treated, and guide the clinician to accurately navigate the instruments to the correct location [4].

Registration is a critical component of image-guided surgery, as it enables accurate "blending" of all disparate pieces of information (pre- and intra-operative data, the patient, and surgical tools) into a common framework. Due to their mathematical complexity and computational inefficiency, some registration algorithms may not be suitable for use in time-critical interventional applications in the operating room (OR). Instead, fast, simple, and OR-friendly registration techniques are employed, often at the expense of misalignments (Fig. 1) present in the visualization environment [5].

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 361–368, 2009.

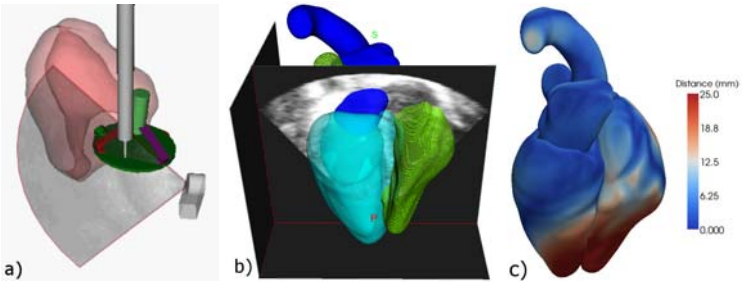


Fig. 1. a) AR environment used to guide an *in vivo* off-pump procedure in a porcine subject, showing pre-operatively generated anatomical model, tracked 2D US image, and virtual tool models; b) Pre-operative model registered to intra-operative US image, and c) error map showing misalignments across the surface model

Recently, Ma *et al.* [6] proposed a feature-based registration technique that relies on the alignment of the left ventricular surface and centerline of the descending aorta to fuse pre- and intra-operative data using a weighted iterative closest point (ICP) registration approach; similarly, we have shown clinically-suitable fusion of pre-operative models and intra-operative US data via aligning reconstructed valve annuli *et al.* [7]. While the features driving the registration are different, both techniques provide comparable anatomical alignment (4-5 mm) of the pre- and intra-operative data. However, although clinically favourable, the achieved alignment may not be suitable for model-guided therapy, without "refined guidance" provided via real-time intra-operative imaging.

We employ an augmented reality (AR)-assisted image guidance platform [8] that integrates real-time trans-esophageal echocardiography (TEE), enhanced with pre-operative heart models, and surgical instrument localization. This platform allows clinicians to explore the intracardiac environment using the pre-operative models as guides for tool navigation [9], but therapeutic success may be subject to the model-to-subject registration accuracy.

The goal of this work is to mimic *in vitro* model-to-subject misregistrations similar to those encountered in cardiac applications and show how our guidance platform provides sufficient navigation information to maintain accurate targeting, in spite of slight misregistrations. This work demonstrates the value of pre-operative models as a means of quickly navigating to the neighbourhood of the target, with real-time US used to refine the targeting operation.

2 Materials and Methods

We mimicked *in vitro* endocardial ablation procedures in a beating heart phantom. To demonstrate the robustness of our surgical platform, we then altered the position and orientation of the pre-operative phantom model in the visualization environment with respect to its real-world counterpart. This procedure enabled us to simulate misalignments similar to those observed in clinical applications and quantify the success of "therapy delivery" to specific endocardial targets.

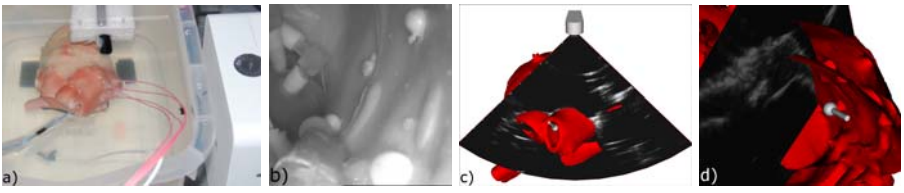


Fig. 2. a) Beating heart phantom setup, with catheter and US probe; b) Endoscopic view of the endocardial surgical targets; c) Virtual surgical field (phantom model, tracked 2D US probe and fan), accompanied by a cut-away endocardial view showing targets and virtual catheter tip

2.1 Experimental Design

The study was performed on a pneumatically-actuated beating heart phantom (The Chamberlain Group, Great Barrington, MA, USA). Ten CT-visible fiducials were attached onto the epicardial surface of the phantom to assist with the model-to-phantom registration. Four surgical targets (3.2 mm Teflon spheres) were implanted into the endocardial surface, mimicking right atrial pathology. Their position was tracked throughout the cardiac cycle using 5 degree-of-freedom NDI Aurora magnetic sensors embedded within each sphere. Therapy delivery consisted of users navigating the tip of a tracked catheter to each target (Fig. 2).

2.2 Image-Guidance Platform

Model-guided Visualization. In a typical procedure, a pre-operative image or model of the patient's heart featuring the surgical targets is registered to the patient in the OR. To mimic the clinical work flow, we first acquired a gated (60 bpm) cine CT dataset (0.48 x 0.48 x 1.25 mm^3) that depicts the phantom at 20 phases over the cardiac cycle. Using segmentation tools available in the Vascular Modeling ToolKit (http://www.vmtk.org), we reconstructed virtual surface models at each cardiac phase and rendered them in "cine" mode for dynamic visualization (Fig. 2).

Real-time Imaging. TEE is extensively used in interventional guidance thanks to its excellent real-time capabilities and OR compatibility. A distinct feature of our surgical platform consists of the acquisition of *tracked* 2D TEE images in real-time, enabling their display within anatomical context provided by the virtual model and relative to the tracked surgical tools. While most of our work is based on tracked TEE for intra-procedural imaging, similar results can be achieved using intracardiac or laparoscopic US imaging as shown in [10]. Here the TEE transducer is tracked using a 6 DOF magnetic sensor attached to the probe and calibrated using a Z-string technique [11].

Virtual-to-Real World Registration. The key to building accurate navigation environments lies within the registration of all components into a common

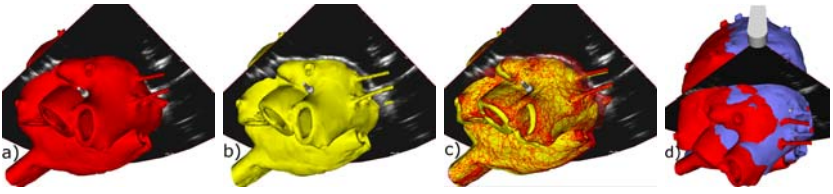


Fig. 3. a) Well-registered and (b) misregistered phantom model w.r.t. its physical counterpart (note epicardial contour in US image); c) Superimposed well-aligned (red wireframe) and translated (yellow surface) models, and d) Superimposed well-registered (red) and rotated (blue) model, mimicking clinically encountered misalignments

framework. The virtual models of surgical tools, US transducer and echo images are intrinsically registered to the tracking coordinate system and to each other via their respective "tool-to-sensor" calibration transforms.

The challenging step is registering the patient (in this case the heart phantom), initially in tracking coordinate system, to that of the pre-operative model. In the clinic this step is achieved via either landmark- or feature-based registration, and resulting misalignments further propagate and impact targeting accuracy. We used a point-based registration algorithm involving epicardial fiducial markers to register the virtual model to the physical phantom. Temporal alignment was achieved by synchronizing the "model heart rate" with the ECG signal driving the actuator, resulting in nearly real-time visualization.

Inducing "Model-to-Subject" Misregistrations. Most image-guided cardiac procedures, including those performed on the open heart, suffer from misalignments between the pre- and intra-operative data. Surgical target locations, although accurately labeled pre-operatively, may not align with their intra-operative location, and may even show as situated outside the cardiac chamber, in which case careless navigation could lead to severe outcomes (Fig. 3).

We simulated two misalignment scenarios often encountered in the OR. The former mimics a translational misalignment, similar to that observed after the pericardial sac is opened to access the heart; the latter mimics a rotational misalignment, arising due to a slightly different orientation of the heart between imaging and intervention. For this study, the phantom model was mapped to its "new" location using the misalignment transforms explored above. Hence, the position and orientation of all tracked tools and US probe remained the same with respect to one another and to the physical phantom, but their location with respect to the model displayed in the visualization environment was changed.

2.3 Surgical Guidance

To evaluate the effect of model-to-subject misregistration on targeting accuracy, we conducted several experiments where users relied solely on model-guided visualization or VR-enhanced US guidance for therapy delivery, under both well-aligned, as well as misaligned conditions. In a previous study, the same

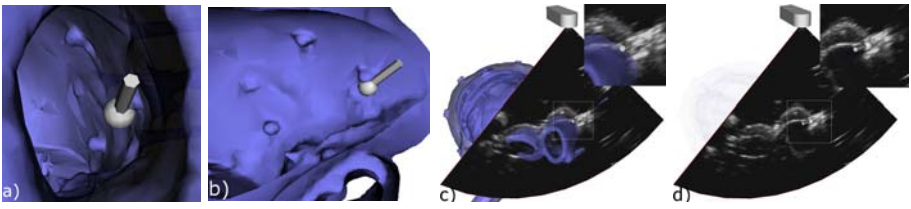


Fig. 4. US enhanced model-assisted guidance work flow: a) initial navigation via virtual anatomy and tool model using two orthogonal views (b), followed by US model enhancement (c), and final target identification and tool tip positioning performed under real-time US guidance (d)

experiment was performed under US image guidance alone to illustrate the limitations of 2D imaging regarding tool navigation. Moreover, to establish a baseline (i.e., positive control) with respect to targeting accuracy, therapy delivery was also conducted under endoscopic guidance. While not suitable for visualization in blood-filled cavities, endoscopic guidance closely mimics a direct vision procedure and its outcomes.

Users were blinded to whether or not the model was properly registered to the phantom, so relying on the model alone might not have led to accurate targeting. Therefore, they adopted a "therapy delivery work flow" where the initial tool to target navigation was performed under model-assisted guidance. Once on target, the display of the model was dimmed, while the tracked 2D US image was emphasized, allowing users to identify the true target location and refine their tool location based on the virtual tool representation and the real-time US image (Fig. 4).

3 Evaluation and Results

Three novice users conducted the *in vitro* catheter navigation on four surgical targets, whose positions were tracked simultaneously. The four targets were approached in sequence, and the order of the targets was arbitrarily generated for each sequence. Each user attempted to localize the targets in 4 trials under virtual model guidance alone (VR), as well as US-enhanced model guidance (VR-US), with no prior knowledge of the model-to-phantom registration. In a previous study, the same three users performed the similar task under endoscopic and US image guidance alone. Prior to acquiring the measurements, each user was allowed a short training period to become accustomed to the "surgical navigation" technique. The procedure outcome was assessed according to targeting accuracy — the distance between catheter tip and surgical target when in contact (**Table 1**).

We performed an analysis of variance (ANOVA) followed by a Tukey's post-hoc test (GraphPad Prism 4.0) to identify significant differences in procedure accuracy with respect to the guidance modality employed and to evaluate the effect of the model-to-phantom registration on targeting accuracy.

Table 1. RMS targeting accuracy (mm) w.r.t. guidance modality and model alignment

Guidance Modality	Registered Model				Translated Model		Rotated Model	
	ENDO	US	VR	VR-US	VR	VR-US	VR	VR-US
Global	0.5	14.8	0.9	0.7	2.9	1.1	3.4	1.4
Target 1	0.4	13.4	0.8	0.8	3.2	0.9	3.1	1.1
Target 2	0.3	14.8	1.0	0.8	2.4	1.2	3.2	1.7
Target 3	0.5	20.3	0.7	0.7	2.9	1.4	3.4	1.6
Target 4	0.6	9.4	1.1	0.5	3.1	0.8	3.7	0.9

4 Discussion

Our study following accurate model-to-phantom registration [12] showed that both the model-assisted (VR) and US-enhanced model-assisted (VR-US) guidance led to significantly more accurate targeting ($p < 0.001$) than 2D US imaging alone; moreover, no significant difference was observed between the endoscopic and the two model-assisted guidance modalities ($p > 0.05$) (Fig. 5). Model-assisted guidance also reduced the "procedure" time in half [12].

Next, we analyzed the effect of model-to-subject registration on the guidance modality employed. The chosen "misalignment transforms" induced a surgical target "shift" of 3-5 mm depending on their locations. Following model-assisted guidance, we observed a significant decrease in targeting accuracy ($p < 0.001$) compared to that recorded under well-aligned conditions (Fig. 5). On the other hand, while targeting accuracy was inadequate, targeting precision was maintained; by mapping the "targeted sites" using the inverse of the misalignment transforms, we were able to reconstruct the "true" locations of the surgical targets. Hence, these navigation errors were induced by the misleading environment.

However, once model-assisted guidance was augmented with real-time US, we observed a significant retrieval of targeting accuracy. For the translational misalignment case, targeting accuracy achieved under model-assisted guidance alone dropped to an overall 3 mm RMS; however, it was restored to an overall 1.1

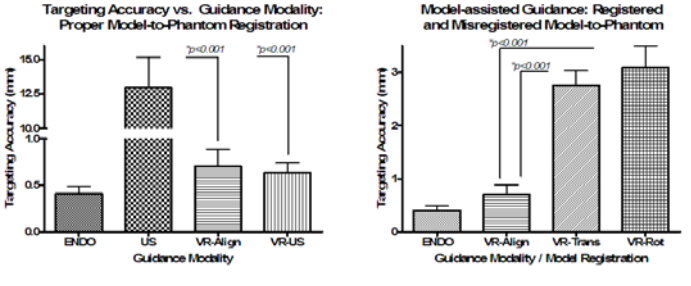


Fig. 5. Superior targeting accuracy is achieved under AR-assisted vs. US imaging alone; however, accuracy of *model-only guidance* (VR) is greatly reduced under misalignments

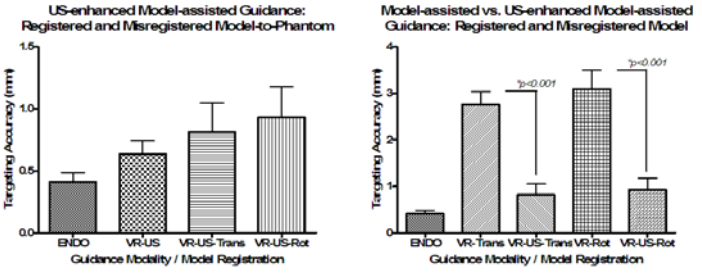


Fig. 6. US-enhanced model-assisted guidance (VR-US) shows consistent targeting accuracy independently of model-to-subject registration. Note the significantly improved accuracy with US-enhanced model-assisted (VR-US) guidance under misregistration.

mm RMS error via US-enhanced guidance ($p < 0.001$). Similarly, model-assisted guidance accuracy was significantly improved with the addition of US imaging, from 3.1 mm targeting error to 1.4 mm RMS (Fig. 5). Moreover, virtual reality-enhanced US guidance maintained a high level of accuracy regardless of the model-to-subject registration ($p < 0.001$) (Fig. 6).

For qualitative evaluation, we recorded "targeting maps" after each "therapy delivery" session. A compact distribution of targeted sites was observed under endoscopic guidance, and maintained under both model-assisted and US-enhanced model-assisted guidance. The main drawbacks of US image guidance arose due to limited information provided for navigation; the lack of context and poor instrument perception, 2D US alone cannot adequately portray the 3D surgical environment. US image guidance alone hampers the navigation step, leading to large targeting errors, while model-assisted guidance alone hampers the positioning step, leading to precise, but insufficiently accurate targeting.

5 Conclusions

This study is key for evaluating our AR-assisted surgical guidance platform. We have shown that despite slight misalignments between the virtual model and the subject, real-time US does provide sufficient information to identify the true target location and compensate for the navigation errors, ultimately enabling consistent targeting accuracy on the order of 1-1.5 mm.

Acknowledgments. The authors thank Mathew Carias, Carling Cheung, Danielle Pace, Petar Seslija and Dr. Aaron So for help with data acquisition. In addition, we acknowledge funding for this work provided by the following Canadian agencies: NSERC, CIHR, HSFC, ORDCF, OIT and CFI.

References

1. Doty, D.B., Flores, J.H., Doty, J.R.: Cardiac valve operations using a partial sternotomy technique. J. Card. Surg. 15, 35–42 (2000)
2. Vassiliades, T.A., Block, P.C., Cohn, L.H., et al.: The clinical development of percutaneous heart valve technology. J. Thorac. Cardiovasc. Surg. 129, 970 (2005)
3. Lutter, G., Ardehali, R., Cremer, J., et al.: Percutaneous valve replacement: current state and future prospects. Ann. Thor. Surg. 78, 2199–2206 (2004)
4. Rettmann, M.E., Holmes, D.R., Robb, R.A.: An integrated system for real-time image-guided cardiac catheter ablation. In: Proc. of Medicine Meets Virtual Reality. Stud. Health Technol. Inform., vol. 119, pp. 455–460 (2006)
5. Peters, T.M., Cleary, K.: Image-guided Interventions: Technology and Applications. Springer, Heidelberg (2008)
6. Ma, Y., Rhode, K.S., Penney, G.P., et al.: Echocardiography to magnetic resonance image registration for use in image-guided electrophysiology procedures. In: Proc. of SPIE - Medical Imaging, vol. 7261, 72610Q–8 (2009)
7. Linte, C.A., Moore, J., Wedlake, C., et al.: Inside the beating heart: An in vivo feasibility study on fusing pre- and intra-operative imaging for minimally invasive therapy. Int. J. CARS 4, 113–122 (2009)
8. Linte, C.A., Moore, J., Wiles, A.D., et al.: VR-enhanced US guidance: A novel technique for intracardiac interventions. Comput. Aided Surg. 13, 82–94 (2008)
9. Sauer, F.: Image registration: Enabling technology for image-guided surgery and therapy. Proc. of IEEE Eng. Med. Biol. Soc., 7242–7245 (2005)
10. Wiles, A.D., Moore, J., Wedlake, C., et al.: Object identification accuracy under ultrasound enhanced virtual reality for minimally invasive cardiac surgery. In: Proc. of SPIE - Medical Imaging, vol. 6918, 69180E–12 (2008)
11. Gobbi, D.G., Comeau, R.M., Peters, T.M.: Ultrasound probe tracking for real-time ultrasound/MRI overlay and visualization of brain shift. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 920–927. Springer, Heidelberg (1999)
12. Linte, C.A., Moore, J., Wiles, A.D., et al.: In vitro cardiac catheter navigation via augmented reality surgical guidance. In: Proc. of SPIE. - Medical Imaging, vol. 7261, 72610–9 (2009)

Patient Specific 4D Coronary Models from ECG-gated CTA Data for Intra-operative Dynamic Alignment of CTA with X-ray Images*

Coert T. Metz[1], Michiel Schaap[1], Stefan Klein[1], Lisan A. Neefjes[2], Ermanno Capuano[2], Carl Schultz[2], Robert Jan van Geuns[2], Patrick W. Serruys[3], Theo van Walsum[1], and Wiro J. Niessen[1,4]

[1] Dept. of Medical Informatics and Radiology, Erasmus MC, Rotterdam, The Netherlands
c.metz@erasmusmc.nl
[2] Dept. of Radiology and Cardiology, Erasmus MC, Rotterdam, The Netherlands
[3] Dept. of Cardiology, Erasmus MC, Rotterdam, The Netherlands
[4] Faculty of Applied Sciences, Delft University of Technology, The Netherlands

Abstract. We present an approach to derive patient specific coronary models from ECG-gated CTA data and their application for the alignment of CTA with mono-plane X-ray imaging during interventional cardiology. A 4D (3D+t) deformation model of the coronary arteries is derived by (i) extraction of a 3D coronary model at an appropriate cardiac phase and (ii) non-rigid registration of the CTA images at different ECG phases to obtain a deformation model. The resulting 4D coronary model is aligned with the X-ray data using a novel 2D+t/3D+t registration approach. Model consistency and accuracy is evaluated using manually annotated coronary centerlines at systole and diastole as reference. Improvement of registration robustness by using the 2D+t/3D+t registration is successfully demonstrated by comparison of the actual X-ray cardiac phase with the automatically determined best matching phase in the 4D coronary model.

1 Introduction

Coronary artery disease is one of the main causes of death in the world [1]. Atherosclerotic plaque build-up in the coronary arteries leads to vessel narrowing and occlusion, which may eventually cause heart failure. An often applied therapy to reopen occluded vessels is coronary angioplasty. During these procedures X-ray imaging is applied to visualize the arteries and guide wires. This modality unfortunately lacks depth information due to its projective nature, and only visualizes the vessel lumen, which makes the visualization of total occlusions impossible. Coronary CTA, on the contrary, also allows the visualization of coronary plaque and total occlusions. Integration of this information into the

* This work was supported by the Dutch Foundation for Scientific Research, NWO, STW.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 369–376, 2009.

interventional scene may thus improve the treatment of these disorders. This is, for example, especially important when an ablating catheter is needed to open occlusions, making correct positioning of the catheter essential to prevent perforation of the vessel wall.

The goal of our work is therefore to integrate pre-operative CTA data into the interventional scene. Due to cardiac motion, the coronary arteries are under continuous deformation during therapeutic procedures. A rigid registration approach is therefore not sufficient to align the CTA data with the intra-operative X-ray images. Moreover, diagnostic CTA scans are generally reconstructed at those phases of the cardiac cycle where the deformation of the coronary arteries is minimal. In other phases, image quality is not only degraded by more excessive motion of the coronary arteries, but also due to lowering the tube current by ECG-derived pulsing windows which aim to minimize effective patient dose.

In this work, we employ the phase with minimal coronary motion for 3D coronary reconstruction, and use the lower quality phases for defining a deformation model. Together, this yields a 4D (3D+t) coronary model, which can be applied to make high quality information available over the complete cardiac cycle. This information can be used to improve the robustness of CTA/X-ray alignment, which is useful for improved navigation in interventional cardiology. To this end, we propose a novel 2D+t/3D+t registration approach.

Related work with respect to the 4D model building has, for example, been presented by Wierzbicki et al., who present an approach to derive heart motion from CTA and MRA data by means of a non-rigid registration procedure [2]. To the best of our knowledge, the application of this approach to derive patient specific 4D coronary models has not been reported previously. Moreover, we show the application of these models for the registration of CTA data to intra-operative X-ray data; this 2D+t/3D+t registration approach is also novel.

Static registration of simulated X-ray images and CTA data has been performed by Turgeon et al., who presented results on both mono- and biplane registration at end-diastole [3]. Deformable 2D/3D registration of vasculature was presented by Zikic et al., who constrained the deformation of the vessels based on a priori information about plausible deformations [4]. Especially for mono-plane X-ray imaging, these free-form deformations can in theory introduce ambiguity, *i.e.* multiple 3D deformations can result in similar projection images. The approach presented in this work is different in two aspects: (i) the movement of the 3D vasculature is restricted to patient specific deformations derived from ECG-gated CTA data and (ii) a window of multiple sequential X-ray images is simultaneously aligned to a window of multiple sequential 4D coronary model positions.

2 Method

The method consists of a number of steps. First, A 3D centerline or lumen model of the coronary arteries is derived manually or semi-automatically from a phase in the cardiac cycle with minimal motion. Deformation of the coronary arteries is

subsequently derived from the 4D CTA data by means of a non-rigid registration procedure. We assume that the coronary arteries move with the larger structures in the image to which they are attached, such as the heart chambers, which is especially advantageous in those phases of the cardiac cycle where the arteries are not well visible due to motion artifacts. Finally, the resulting 4D coronary model is applied in a new 2D+t/3D+t registration procedure for the dynamic alignment of CTA with X-ray data. The overall method is depicted in Fig. 1; details of the different steps are presented below.

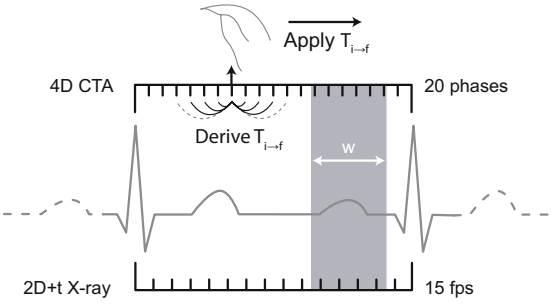


Fig. 1. Overview of 4D model creation and 2D+t/3D+t registration approach

2.1 Deriving the 4D Coronary Model

The registration approach applied to derive the coronary deformation from the 4D CTA data is similar to the one proposed by Wierzbicki et al. [2]. The 4D datasets are divided into n different 3D images, denoted I_i, where $i \in \{0 \dots n-1\}$ indicates the phase in the cardiac cycle, further referred to as time-point. A 3D model of the coronary arteries is derived manually or semi-automatically from one of these images, *e.g.* at end-systole or end-diastole where image quality is in general higher than in other parts of the cardiac cycle. This 3D coronary centerline/lumen model is subsequently mapped to the other time-points by applying deformations that are derived by a non-rigid registration procedure. The image at the time-point in which the 3D model is defined is chosen as reference image (I_f) as we want to minimize the error of the registration to this time-point in order to find the model deformation as accurate as possible. All other time-point images $I_{i\neq f}$ are subsequently registered to this fixed image. Registration starts using I_{f+1} as moving image and the resulting transformation T_{f+1} is used as the initial transformation for registration of the next time-point image I_{f+2}.

Registration is performed using a multi-resolution approach in which Gaussian filtering is applied to reduce the scale of the image in the resolution steps. The optimal B-spline deformation is searched using a stochastic gradient descent optimization approach in which mutual information is used as similarity measure [5,6].

Using the transformations $T_{i\neq f\to f}$ resulting from the registration approach, the 3D coronary model is deformed to all time-points of the 4D CTA data. As these time-points do not necessarily match the frame rate of the X-ray data

acquired during the intervention, spline interpolation is applied to derive intermediate coronary model positions.

2.2 2D+t/3D+t Registration Procedure

For the alignment of the CTA with the X-ray data, a rigid intensity-based registration framework is applied [7]. Instead of matching one time-point of the 4D coronary model with one X-ray image, we propose to match multiple sequential time-points of the 4D coronary model in a window w to multiple sequential X-ray images simultaneously. This window includes the time-point we want to match and the $w-1$ previous time-point positions, mimicking the clinical situation. This approach is expected to improve the robustness of the registration procedure as there is additional information in the time-dimension of the data, which is especially useful in the case of mono-plane X-ray imaging.

On the X-ray series, static structures are removed by subtracting from each pixel the average pixel intensity of this pixel over time. Before registration, the 4D model of the coronary arteries is automatically positioned with its center of mass on the isocentre of the X-ray coordinate system. Manual initialization is subsequently performed by applying one 2D translation in the plane orthogonal to the projection line for the complete cardiac cycle. Optimization is performed using a multi-resolution gradient ascent procedure. For every pose estimation, digitally reconstructed radiographs (DRRs) of the coronary lumen model are rendered on the graphical processing unit and compared with the real X-ray images using normalized cross correlation (NCC). The resulting NCC values are averaged over the time window w to derive the quality of the current pose.

3 Experiments

3.1 CTA and X-ray Data

CT coronary angiography for thirty-one patients with known or suspected coronary artery disease was acquired using a dual-source CT system (Somatom Definition; Siemens Medical Solutions, Forchheim, Germany). No beta-blockers were administered in preparation for the scan. The entire volume of the heart was scanned in one breath-hold, while simultaneously recording the ECG trace. An optimized heart-rate-dependent ECG pulsing protocol was applied. Outside the ECG pulsing window a tube current of 4% was used in order to reduce the effective dose of the CT examination. Multi-phase 4D datasets were reconstructed at 5% steps from 0% to 95% of the RR-interval, with a slice thickness of 3 mm and slice increment of 1.5 mm. Voxelsizes of the reconstructed volume are approximately 0.7x0.7x1.5 mm^3.

X-ray data was acquired with an AXIOM-Artis C-arm system (Siemens Medical Solutions, Forchheim, Germany). The pixel size of these images is 0.22x0.22 mm^2.

3.2 4D Coronary Model Consistency and Accuracy

The 4D coronary models were evaluated using manually annotated coronary centerlines at both systole and diastole of the cardiac cycle. In every dataset one of the following vessels was randomly selected for annotation: right coronary artery (RCA), left anterior descending artery (LAD) or left circumflex artery (LCX). In total 11 LADs, 11 LCXs and 9 RCAs were annotated. Two experiments were conducted: one to assess model consistency and one to assess model accuracy.

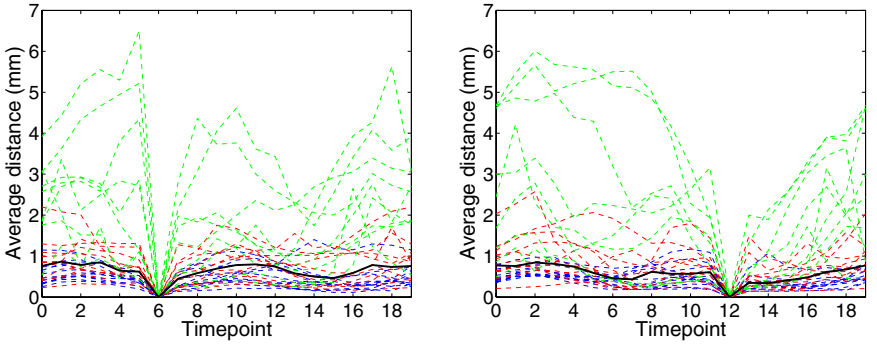
Consistency was measured as follows. First, coronary deformation was derived using the approach described in section 2.1. This deformation was also derived by applying the registration procedure in reverse order, *viz.* by registering image I_{f-1} to I_f, followed by a registration of I_{f-2} to I_f using $T_{f-1 \rightarrow f}$ as initial deformation, etc. The two 4D coronary centerline models are compared by measuring the average distance between the resulting centerlines at every time-point, which is defined as the area between the centerlines divided by the length of the annotated centerline. This experiment is performed using both the image at systole and diastole as reference image in the registration procedure.

The accuracy of the 4D coronary models was determined using the same centerlines as used for the consistency experiment. By applying the deformations resulting from the registration procedure to the centerlines annotated at systole, a deformed version of these centerlines at diastole is derived and compared to the actually annotated centerline at this time-point. Centerlines are clipped at their distal parts, as it is not always possible to annotate them equally distal in both images, due to low resolution of the data and motion artifacts. Clipping is performed by determining the Euclidian closest points on both centerlines and disregarding all points distal from these points in the evaluation. Accuracy is measured by computation of the average distance between the clipped centerlines. The same procedure is applied for the centerlines annotated at diastole.

3.3 2D+t/3D+t Registration

To demonstrate an improvement in registration robustness using the proposed 2D+t/3D+t registration approach, a 4D coronary lumen model was constructed for one dataset for which X-ray data was available. This was performed by manual extraction of the centerlines of the vessels at diastole, followed by automatic segmentation of the lumen [8], and subsequent manual correction of the segmentation results.

One RR-interval of the X-ray data was selected and the 4D coronary lumen model was matched using a 2D/3D+t approach ($w = 1$) and the proposed 2D+t/3D+t approach ($w = 4$). For both approaches and for each X-ray image in the RR-interval, we search exhaustively for the best matching time-point position in the 4D coronary lumen model, by consecutively registering the 3D coronary lumen model at each time-point position to the 2D X-ray image. Subsequently, the best matching time-point is selected as the one yielding the maximum NCC value. In the ideal situation, resulting coronary lumen model time-point positions should exactly match the actual X-ray time-point positions.



(a) Systole (b) Diastole

Fig. 2. Model consistency: the average distance between the centerlines is plotted against the time-point index number. Each line represents a dataset (blue=LAD, red=LCX and green=RCA). The median over all centerlines is plotted in black.

4 Results and Discussion

4.1 4D Coronary Model Consistency and Accuracy

Graphs showing the results for the model consistency experiment can be found in Fig. 2. Results for both using the image at systole and diastole as reference image in the registration procedure are presented. The average distance between the centerlines is plotted against the time-point index number. Each line in the plot represents a dataset. The thick black line represents the median distance over all datasets, which is below 1 mm for both the systolic and diastolic experiment. Model accuracy is presented in Fig. 3, which displays a boxplot of the distance between the centerlines at systole and diastole before (I) and after (R) registration, and the difference between these two (D). Distances are averaged over systolic and diastolic measurements. Results are presented for all vessels combined and for the LAD, LCX and RCA separately. It can be noticed that the LAD shows the least deformation during the cardiac cycle and the RCA the most. The median accuracy is approximately 1 mm.

4.2 2D+t/3D+t Registration

Results of the 2D+t/3D+t experiment are shown in Fig. 4. X-ray time-point positions are plotted against the best matching CT time-point. The dashed line shows the results for $w = 1$ and the solid line for $w = 4$. It can be noticed that using a larger window in the registration procedure results in a much better time-point position derivation than the 2D/3D+t procedure. For some time-phases at the end of the RR-interval, registration ended up in a local optimum, caused by using the same manual initialization for the complete cardiac cycle, which is for some phases outside the capture range of the registration procedure. This may be solved in the future by using the resulting optimal pose for the previous

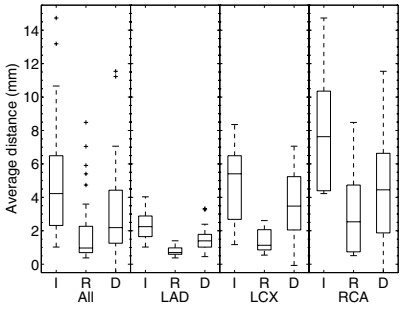


Fig. 3. Accuracy of 4D coronary models. (I)nitial distance between systole and diastole centerlines, after (R)egistration, and (D)ifference (I-R).

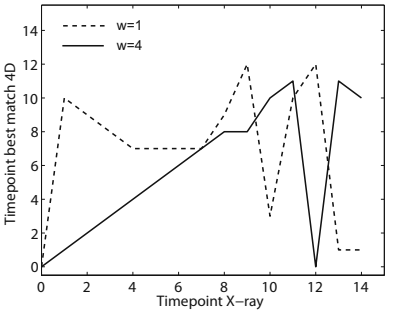


Fig. 4. X-ray time-point positions plotted against best 4D model time-point position resulting from the 2D+t/3D+t registration procedure

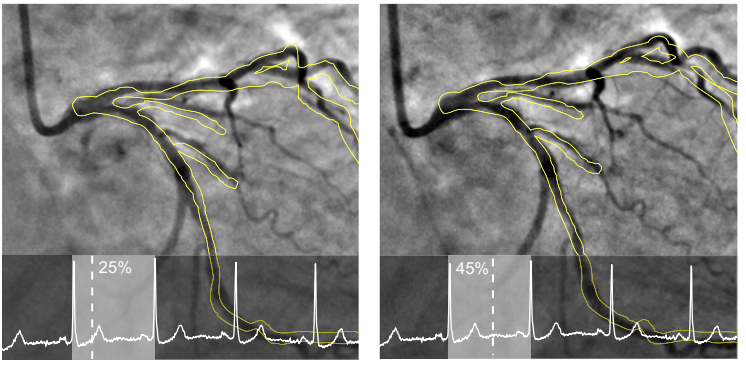


Fig. 5. Examples of 2D+t/3D+t registration results at 20% and 40% of the RR interval. Projected coronary model boundaries are overlayed on the background-subtracted X-ray images

time-point as initial pose for the time-point under consideration. Illustrations with registration results at two phases in the cardiac cycle are shown in Fig. 5. Although this experiment was limited to one dataset, it demonstrates the feasibility of this approach. Moreover, it indicates that registration of CTA and mono-plane X-ray images, which is an ill-posed problem, can benefit from the information in the time dimension of the imaging data. Some misalignment is still visible in Fig. 5, which may be caused by irregular heart deformation, inaccurate reporting of the projection geometry of the X-ray system in the DICOM headers, errors in the 4D coronary model, or non-rigid deformation caused by the breathing cycle. The registration procedure for derivation of the 4D coronary model still needs some improvements, especially to be better able to handle the more excessive motion of the RCA. It can, for example, be advantageous to take the cyclic behavior of the heart motion into account or to imply a smoothness constraint on the registration in the time-dimension of the image. Furthermore,

as we discretized the 4D coronary model in the time dimension, there may still exist a small time mismatch between the X-ray time-point and coronary model time-point.

5 Conclusions

We presented a method to derive patient specific 4D coronary models from ECG-gated CTA data and demonstrated the usefulness of these models for better relating pre-operative CTA to intra-operative X-ray images, which can improve guidance during interventional cardiology procedures. Hereto, a novel 2D+t/3D+t registration procedure was developed and we demonstrated that this procedure was better able to determine the correct CTA time-point position for the X-ray phase of interest, which is an indication that registration robustness can be improved compared to a 2D/3D+t approach. We conclude that the combination of the 4D coronary models and the 2D+t/3D+t registration method is a promising approach for improved image guidance in interventional cardiology.

References

1. Rosamond, W., Flegal, K., Furie, K., et al.: Heart disease and stroke statistics – 2008 update: Report from the american heart association statistics committee and stroke statistics subcommittee. Circulation 117, 25–146 (2008)
2. Wierzbicki, M., Drangova, M., Guiraudon, G., Peters, T.: Validation of dynamic heart models obtained using non-linear registration for virtual reality training, planning, and guidance of minimally invasive cardiac surgeries. Med. Image Anal. 8(3), 387–401 (2004)
3. Turgeon, G.-A., Lehmann, G., Guiraudon, G., Drangova, M., Holdsworth, D., Peters, T.: 2D-3D registration of coronary angiograms for cardiac procedure planning and guidance. Med. Phys. 32(12), 3737–3749 (2005)
4. Zikic, D., Groher, M., Khamene, A., Navab, N.: Deformable registration of 3D vessel structures to a single projection image. In: Proc. SPIE (Medical Imaging: Image Processing) (2008)
5. Klein, S., Staring, M., Pluim, J.P.W.: Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. IEEE Transactions on Image Processing 16(12), 2879–2890 (2007)
6. Mattes, D., Haynor, D.R., Vesselle, H., Lewellen, T.K., Eubank, W.: PET-CT image registration in the chest using free-form deformations. IEEE Trans. Med. Imag. 22(1), 120–128 (2003)
7. Penney, G.P., Batchelor, P.G., Hill, D.L., Hawkes, D.J., Weese, J.: Validation of a two- to three-dimensional registration algorithm for aligning preoperative CT images and intraoperative fluoroscopy images. Med. Phys. 28(6), 1024–1032 (2001)
8. Schaap, M., Neefjes, L., Metz, C., et al.: Coronary lumen segmentation using graph cuts and robust kernel regression. In: Information Processing in Medical Imaging (2009)

Towards Interactive Planning of Coil Embolization in Brain Aneurysms

Jeremie Dequidt[1], Christian Duriez[1], Stephane Cotin[1], and Erwan Kerrien[2]

[1] INRIA Lille Nord Europe - Alcove Team
50, avenue Halley 59655 Villeneuve d'Ascq, France
{jeremie.dequidt,christian.duriez,stephane.cotin}@lifl.fr
[2] INRIA Nancy Grand Est - Magrit Team
Campus Scientifique, 54506 Vandoeuvre-lès-Nancy, France
erwan.kerrien@loria.fr

Abstract. Many vascular pathologies can now be treated in a minimally invasive way thanks to interventional radiology. Instead of open surgery, it allows to reach the lesion of the arteries with therapeutic devices through a catheter. As a particular case, intracranial aneurysms are treated by filling the localized widening of the artery with a set of coils to prevent a rupture due to the weakened arterial wall. Considering the location of the lesion, close to the brain, and its very small size, the procedure requires a combination of careful planning and excellent technical skills. An interactive and reliable simulation, adapted to the patient anatomy, would be an interesting tool for helping the interventional neuroradiologist plan and rehearse a coil embolization procedure. This paper describes an original method to perform interactive simulations of coil embolization and proposes a clinical metric to quantitatively measure how the first coil fills the aneurysm. The simulation relies on an accurate reconstruction of the aneurysm anatomy and a real-time model of the coil for which sliding and friction contacts are taken into account. Simulation results are compared to real embolization procedure and exhibit good adequacy.

1 Introduction

Intracranial aneurysms are abnormal bulges of the arterial wall that may occasionally rupture, causing hemorrhage, stroke or death. The coil embolization technique involves approaching the aneurysm by navigating through the vascular network using a catheter until the diseased blood vessel is reached, thus avoiding the need to open the skull. Once the aneurysm is reached, a micro-catheter is then partially inserted inside the aneurysm, and the coil is carefully deployed using a guidewire pushing the coil out of the micro-catheter. Usually, several coils are deployed to fill the aneurysm, allowing to clot and eventually to stop blood from entering the aneurysm.

Although the overall morbidity and mortality associated with endovascular repair is lower than for surgery (9.5% at one year, compared to 12.2%) the

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 377–385, 2009.

technical challenges associated with coil embolization remain important. It was however shown recently that the risk of complications with coil embolization of unruptured aneurysms decreases dramatically with physician experience [1]. Indeed, complete embolization is usually determined by subjective assessments of radiologists. Although several objective criteria have recently been introduced to help radiologists during the procedure [2] simulation systems can offer more functionalities and can be used for physician training as well as for pre-operative planning.

Previous work in the area of real-time or near real-time simulation for interventional radiology has mainly focused on training. For instance, Li *et al.* [3], Hoefer *et al.* [4], Alderliesten *et al.* [5], or Cotin *et al.* [6,7] have proposed different approaches for modeling either catheter deformation and more generally catheter navigation in vascular networks. Regarding the planning of interventional radiology procedures, the most representative work was in the area of stent deployment. Laroche *et al.*[8] proposed a finite element model for a patient-specific planning of balloon angioplasty and stent implantation. Complex device-artery interactions occurring during stent deployment were computed, however the simulation was not performed in real-time.

The objective of the work described in this paper is to provide a tool that can be used for planning the procedure. This involves selecting the proper characteristics of the coil based on the patient anatomy, but also interactively deploying the virtual coil(s) to plan or rehearse the procedure. The interactive aspect in this planning plays a key role as the interventional radiologist constantly controls the motion of the coil and as such influences the final coil placement within the aneurysm.

Section (2) presents the different models used for the simulation: the reconstruction of vascular networks based on rotational angiography and the coil model based on beam elements. The contact modeling to tackle interactions between the aneurysm and the coils as well as self-collision of the coils are described in section (3) which also introduces the optimizations allowing interactive simulations of coil embolization procedure. Finally, simulation results are presented in section (4) where we introduce a clinical metric in order to compare the simulation to actual coil deployments in patient-specific aneurysms.

2 Models

A complete coil embolization consists in coils with specific rest shapes conforming to the geometry of an aneurysm. To reproduce this matching problem within a simulation, accurate models are required. This section describes the modelling stage of our framework which aims at virtually reproducing the most important features of real data (such as geometry for the vascular network or mechanical properties for the coils).

Patient-Specific Aneurysm Model. The input of our algorithm is a volumetric 3D Xray (3DXA) angiography of the brain. Such 3D images are recognized

as being of a daily clinical usefulness for the planning and follow-up of the treatment of cerebral pathologies [9]. In order to use deformable active surfaces to segment the aneurysm, an initial surface has to be defined. This region of interest is obtained by thresholding the volumetric data with an appropriate value given by the user. The triangulated initial surface is the boundary of this region and is classically obtained using a marching cube algorithm [10]. This first guess is obviously not accurate due to the use of a global threshold. Deformable triangulated surfaces are then used to recover a precise shape of the aneurysm starting from the initial shape.

The method used for the 3D reconstruction of the aneurysm shape from the initial guess V originates in [11]. Let v_i be the vertices of the triangulated mesh. The shape of the aneurysm is defined as the set of vertices V which minimizes the energy term $E = \sum_{i=1}^{n} E_{internal}(v_i) + \lambda E_{external}(v_i)$ where $E_{internal}(v_i) = |v_i - \bar{v_i}|$ imposes smoothness constraints on the surface ($\bar{v_i}$ being the average set of vertices connected to v_i) and $E_{external}$ is the force field created by the image gradient: $E_{external}(v_i) = -|\nabla I(v_i)|$ where $I(v_i)$ is the 3DXA intensity at vertex v_i.

Using active surfaces not only improves the segmentation of the aneurysm but also provides a smooth surface representation, presenting better characteristics for further processing during the simulation. In particular, the computation of the interaction between the coil and aneurysm wall requires a smooth surface. We obtain a more regular representation than what can be obtained using a marching cubes algorithm [10], even when applying a post-processing smoothing.

Coil Model. There are different types of detachable coils but most of them have a core made of platinum, and some are coated with another material or a biologically active agent. All types are made of soft platinum wire of less than a millimeter diameter and therefore are very soft. The softness of the platinum allows the coil to conform to the arbitrary shape of an aneurysm.

The deformation model of the coil is based on the recent work of Dequidt *et. al.* [12]. Coil dynamics is modelled using serially linked beam elements:

$$\mathbf{M}\dot{\mathbf{v}} = \mathbf{p} - \mathbb{F}\ (\mathbf{q}, \mathbf{v}, \mathbf{q}_0) + \mathbf{Hf} \tag{1}$$

where $\mathbf{M} \in \mathbb{R}^{(n \times n)}$ gathers the mass and inertia matrices of the beam elements. $\mathbf{q} \in \mathbb{R}^n$ is the vector of generalized coordinates (each node at the extremity of a beam has six degrees of freedom: three of which correspond to the spatial position, and three to the angular position in a global reference frame). The rest position $\mathbf{q}_0$ depends on the family of coil to be simulated: helical shape or 3D shape consituted of omega loops (see figure 1). $\mathbf{v} \in \mathbb{R}^n$ is the vector of velocity. $\mathbb{F}$ represents internal visco-elastic forces of the beams, and $\mathbf{p}$ gathers external forces. $\mathbf{f}$ is the vector of the contact forces with the aneurism wall, and $\mathbf{H}$ gathers the contact directions. To integrate this model we use backward Euler scheme with a unique linearization of $\mathbb{F}$ per time step.

Fig. 1. Example of coils used in our simulations, left: Boston Scientific helical coil GDC 10, right: 3D GDC built with omega loops [13]

3 Simulation of Coil Deployment

3.1 Modeling Contacts with Aneurysm Walls

Simulating coil embolization requires to accurately model contacts that occur between the coil and the wall of the aneurysm. This contact model must account for the stick and slip transitions that take place during the deployment of the coil. The model also includes a compliant behavior of vessel wall that is close to Boussinesq model [14]. For modeling contacts with friction, we use two different laws, that are based on the contact force and on the relative motion between the coil and the aneurysm wall. The contact law is defined along the normal $\mathbf{n}$ and the friction law, along the tangential $(\mathbf{t}, \mathbf{s})$ space of the contact.

The contact model, based on Signorini's law, indicates that there is complementarity between the gaps δ^n and the contact forces f^n along the normal direction, that is:

$$0 \leq \delta^n \perp \mathrm{f}^n \geq 0 \tag{2}$$

With Coulomb's friction law, the contact force lies within a spacial conical region whose height and direction are given by the normal force, giving two complementarity conditions for stick and slip motion:

$$\begin{array}{ll} [\delta^t \quad \delta^s] = \mathbf{0} \Rightarrow \|[\mathrm{f}^t \quad \mathrm{f}^s]\| < \mu \, \|\mathrm{f}^n\| & \text{(stick condition)} \\ [\delta^t \quad \delta^s] \neq \mathbf{0} \Rightarrow \; [\mathrm{f}^t \quad \mathrm{f}^s] = -\mu \, \|\mathrm{f}^n\| \frac{[\delta^t \quad \delta^s]}{\|[\delta^t \quad \delta^s]\|} & \text{(slip condition)} \end{array} \tag{3}$$

Where the vector $[\delta^t \quad \delta^s]$ provides the relative motion in the tangential space and μ represents the friction coefficient.

The obtained complementarity relations could create *singular* events when it changes from one state to an other: For instance, when a collision occurs at instant $t^\star$, the velocity $\mathbf{v}(t^\star)$ of the coil, at that point, changes instantaneously. The acceleration could then be ill-defined and we can observe some quick changes in the dynamics. Each friction contact creates three nonholonomic constraints along the normal and tangential directions. Our approach allows for processing simultaneously multiple friction contacts, including self-contacts on the coil.

3.2 Simulation Steps

The processing of one simulation step begins by solving equation (1) for all forces except contact forces ($\mathbf{f} = 0$). This *free motion* corresponds essentially to

the deformation of the beam elements under gravity and user force input. Once the *free motion* has been computed, collision detection computes the contact points between the coil model and the aneurysm surface and the points of self-collision. When collisions are detected, the *contact response* is computed. This is a complex aspect that influences greatly the overall behavior of the coil model. To describe the mechanical behavior during contact, the mechanical coupling between the different contact points is modeled. This information is provided by evaluating the compliance matrix in the contact space, called $\mathbf{W}$, for both the coil and the aneurysm. Let's consider a contact α on the node i of the coil (with one constraint along the contact normal $\mathbf{n}$ and two along the tangential friction directions $\mathbf{t}, \mathbf{s}$). $\mathbf{H}_\alpha$ is the matrix of the frame $[\mathbf{n}\ \mathbf{t}\ \mathbf{s}]$. The mechanical coupling of this contact with a contact β (with frame $\mathbf{H}_\beta$) on node j can be evaluated with the following 3×3 matrix:

$$\mathbf{W}_{(\alpha,\beta)} = \mathbf{H}_\alpha^T \left(\frac{\mathbf{M}}{h^2} + \frac{d\mathbb{F}}{hd\mathbf{v}} + \frac{d\mathbb{F}}{d\mathbf{q}} \right)^{-1}_{(i,j)} \mathbf{H}_\beta = \mathbf{H}_\alpha^T \mathbf{C}_{(i,j)} \mathbf{H}_\beta \tag{4}$$

where $\mathbf{C}_{(i,j)}$ is the 3×3 sub-matrix of global compliance matrix $\mathbf{C}$ (inverse of tangent matrix) at the rows of node i and the columns of node j. For the aneurysm wall, the formulation of the coupling is simpler:

$$\mathbf{W}_{(\alpha,\beta)} = \frac{g(d_{ij})}{e}\ \mathbf{H}_\alpha^T\ \mathbf{H}_\beta \tag{5}$$

where e is an elasticity parameter that is homogeneous to Young modulus and $g(d_{ij})$ is a Gaussian function of the distance, defined on the surface, between contact points i and j. The Gaussian function allows a fall-off of the coupling with increasing distance between the contact points. This model is close to Boussinesq's approximation which provides a distribution of the normal contact stress from the elasticity of the surface, around a point of contact [14].

The result of the *contact response* consists in finding the friction contact forces that respect Signorini's and Coulomb's laws. Several works [15] or [6] present Gauss-Seidel iterative approaches that solve this problem. The solver needs an evaluation of a global compliance matrix $\mathbf{W}$, which is the sum of the compliances of the coil and the aneurysm wall. It also needs the value of the relative displacement of the contacting points during the free motion $\boldsymbol{\delta}^{\text{free}}$. When the contact forces are found, during the last step, called *contact correction* we compute the motion associated to the contact forces.

4 Validation and Results

The main contribution of this work is not to provide an extensive study on simulated coil deployments but rather to introduce a global approach, from segmentation to simulation of the coil embolization procedure. This approach includes a validation step, for which we propose an adapted clinical metric. Examples provided in the paper serve as preliminary results and exhibit the relevance of the chosen metric.

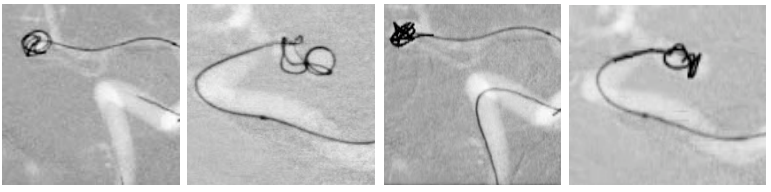


Fig. 2. Examples of our simulation results: (left) real coil embolization (right) our simulated coil embolization with 3D coils

Clinical Metric. As pointed out in the introduction, a successful embolization procedure is determined by the subjective assessment of radiologists and an objective measure known as *Coil Embolization Ratio* (CER) [2]. CER quantifies the amount of coil material that fills up the aneurysm volume. Usually a CER between 20% and 33% is a good value for a complete embolization. The validation proposed in this paper focuses on the deposit of the first coil, which is known to be the most difficult to place and plays a key role in the procedure as it serves as a framework for further embolization. Comparison between final shapes of the real coil and simulated coils is not a suitable solution in this context. Indeed coil deployments are hardly reproducible even with the same initial conditions because they depend of many parameters such as radiologist input, irregularities of the aneurysm surface and can lead to very different final shapes (thus justifying the clinical use of CER for complete embolization). Even experienced interventional radiologists can hardly estimate how the first coil will finally be placed inside the aneurysm. They do however have some criteria to select the first coil [16]: an appropriate size (diameter of the coil has to be close to the aneurysm fundus and larger than the neck), its deployment should cover a large surface of the aneurysm (to serve as a basket in which the subsequent coils will be deposited) and some loops should extend across the neck of the aneurysm. Therefore, we propose in this paper to use a clinical metric derived from the work of Cloft et. al. [16]: the coil embolization surface ratio (CESR) that defines the amount of the aneurysm wall covered by the first coil. The choice of the CESR is twofold: first, it translates into a metric the qualitative criteria proposed by Cloft et. al. and second is less sensitive to small perturbations that could lead to very different shapes. During the beginning of the embolization procedure, the physician constantly evaluates the CESR by watching the 2D fluoroscopic image of the coil deployment. We can obtain easily a measure of CESR and CER of the simulation results but the lack of 3D view of the real coil deployment prevents from having a precise metric on patient data. For a preliminary assessment, we perform the validation of the simulation data by evaluating the CESR on 2D fluoroscopic views as it is widely done by neuroradiologist during procedure.

Simulations. The simulations have been performed using two reconstructed patient aneurysms. For each aneurysm, deployments of a first coil have been simulated. The coil was modeled using 200 beam elements with Young modulus = 75 GPa and Poisson ratio = 0.39. The smooth surface model of the aneurysm

Table 1. CESR measure for the two procedures. Comparable CESR values are obtained when using the same coil as the real procedure. When simulating a helical coil (which is not relevant for this shape of aneurysm), we notice a lower CESR, which is consistent with what would happen in an actual procedure. This shows the predictive nature of our simulation.

	CSER	Error in %
First Patient, experimental data (3D Coil)	81%	0
First Patient Aneurysm with 3D Coil	68%	16
First Patient Aneurysm with Helical Coil	55%	32
Second Patient, experimental data (3D Coil)	24%	0
Second Patient Aneurysm with 3D Coil	17%	30
Second Patient Aneurysm with Helical Coil	11%	54

walls allowed for the simulation of the fast stick and slip transitions that can be observed in actual coil embolizations. Using the shape functions of the beam elements to interpolate the contact points, we obtain a precise auto-collision response. The implicit integration of the coil dynamics allows for a stable behavior of the coil model even during highly constrained transitory motions. The computation time is quite fast at the beginning of the simulation (30 ms for the first loops) but slows down when the contact points become numerous. In order to provide the same initial conditions for the real procedure and the simulated one, the catheter tip is positioned and oriented using a 3D triangulation of 2D orthogonal fluoroscopic images allowing good accuracy (0.2mm) and is less invasive on the deployment than EM trackers. The simulation of the coil is interactive and is controlled using an optical device to acquire the motion induced made by the user. The blood flow resistance has not been taken into account as its influence is said to be negligible by radiologists.

Error Measurement. Our validation tests consist in measuring the CSER metric and computing the relative error with respect to the value obtained from experimental data. The CSER of the experimental data is estimated from 2D images so is a bit overvalued compared to the CSER of our simulations where an accurate area computation of the aneurysm volume covered by coils can be obtained. A summary of the results appears on table 1 and exhibits good adequacy of the behavior of our simulated coils with respect to real ones: indeed our simulated 3D coils provide CSER values close to the experimental values whereas helical coils induce a lower CESR which is consistent with what would happen in an actual procedure.

5 Conclusion and Future Work

In this paper, we have presented a global approach for performing simulation of coil embolization on a patient dataset. The geometric reconstruction of the aneurysm shape is adapted to the simulation requirements. An efficient dynamic coil model is combined with physically based processing of contacts with the

aneurysm wall and the coil itself. The methodology includes a validation process, based on clinical metric. The preliminary results we have obtained are very encouraging, as they illustrate the potential of our approach to capture the key characteristics of this complex procedure and demonstrate a behavior that matches experimental data.

Our next step is to optimize the collision detection and response pipeline in order to obtain real-time performance during a simulation of a whole embolization. We also plan to extend the validation on more patients in order to evaluate the potential use of such a simulation during the planning process of a coil embolization.

Acknowledgments. We would like to thank René Anxionnat, Juan Pablo De la Plata and Frederick Roy for their help on this work.

References

1. Singha, V., Gressa, D., Higashidab, R., Dowdb, C., Halbachb, V., Johnston, S.: The learning curve for coil embolization of unruptured intracranial aneurysms. American Journal of Neuroradiology 23, 768–771 (2002)
2. Satoh, K., Ito, Y., Abe, H.: Measurement of volume ratio to predict coil compaction, on aneurysmal embolization. Interventional Radiology 1(4), 179–182 (1998)
3. Li, Z., Chui, C.-K., Cai, Y., Anderson, J.H., Nowinski, W.L.: Interactive catheter shape modeling in interventional radiology simulation. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, p. 457. Springer, Heidelberg (2001)
4. Hoefer, U., Langen, T., Nziki, J., Zeitler, F., Hesser, J., Mueller, U., Voelker, W., Maenner, R.: Cathi - catheter instruction system. In: Computer Assisted Radiology and Surgery (CARS), Paris, France, pp. 101–106 (2002)
5. Alderliesten, T.: Simulation of Minimally-Invasive Vascular Interventions for Training Purposes. PhD dissertation, Utrecht University (2004)
6. Cotin, S., Duriez, C., Lenoir, J., Neumann, P., Dawson, S.: New approaches to catheter navigation for interventional radiology simulation. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 534–542. Springer, Heidelberg (2005)
7. Dequidt, J., Lenoir, J., Cotin, S.: Interactive contacts resolution using smooth surface representation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 850–857. Springer, Heidelberg (2007)
8. Laroche, D., Delorme, S., Anderson, T., Diraddo, R.: In-vivo validation of a stent implantation numerical model. In: Proc. Medicine Meets Virtual Reality Conference, Studies in Health Technology and Informatics., vol. 125, pp. 265–270 (2007)
9. Anxionnat, R., Bracard, S., Ducrocq, X., Trousset, Y., Launay, L., Kerrien, E.: Intracranial aneurysms: Clinical value of 3d digital subtraction Angiography in the Therapeutic Decision and Endovascular Treatment. Radiology (2001)
10. Lorensen, W., Cline, H.: Marching cubes: A high resolution 3d surface construction algorithm. In: Computer Graphics Proceedings, SIGGRAPH, vol. 21, pp. 163–169 (1987)
11. Lachaud, J.O., Montanvert, A.: Deformable meshes with automated topology changes for coarse-to-fine 3D surface extraction. Medical Image Analysis 3(2), 187–207 (1999)

12. Dequidt, J., Marchal, M., Duriez, C., Kerrien, E., Cotin, S.: Interactive simulation of embolization coils: Modeling and experimental validation. In: Proceedings of MICCAI (2008)
13. Cloft, H.J., Joseph, G.J., Tong, F.C., Goldstein, J.H., Dion, J.E.: Use of three-dimensional guglielmi detachable coils in the treatment of wide-necked cerebral aneurysms. American Journal of Neuroradiology (1999)
14. Pauly, M., Pai, D., Leonidas, G.: Quasi-rigid objects in contact. In: Proceedings of ACM SIGGRAPH Symposium on Computer Animation, pp. 109–119 (2004)
15. Jourdan, F., Alart, P., Jean, M.: A gauss-seidel like algorithm to solve frictional contact problems. Comp. Meth. In: Appl. Mech. and Engin, 33–47 (1998)
16. Cloft, H.J., Joseph, G.J., Tong, F.C., Goldstein, J.H., Dion, J.E.: Use of three-dimensional guglielmi detachable coils in the treatment of wide-necked cerebral aneurysms. American Journal of NeuroRadiology, 1312–1314 (2000)

Temporal Estimation of the 3d Guide-Wire Position Using 2d X-ray Images

Marcel Brückner[1], Frank Deinzer[2], and Joachim Denzler[1]

[1] Chair of Computer Vision, Friedrich Schiller University of Jena, Germany
{marcel.brueckner,joachim.denzler}@uni-jena.de
[2] Siemens AG, Healthcare Sector, Germany and University of Applied Sciences Würzburg-Schweinfurt, Germany
frank.deinzer@fhws.de

Abstract. We present a method for realtime online 3d reconstruction of a guide-wire or catheter using 2d X-ray images, which do not have to be recorded from different viewpoints. No special catheters or sensors are needed. Given a 3d patient data set and the projection parameters, we use recursive probability density propagation to estimate a probability distribution of the current positions of guide-wire parts. Based on this distribution, we extract the optimal guide-wire position using regularization techniques. We describe the guide-wire position by a uniform cubic B-spline. Experiments on simulated and phantom data demonstrate the high accuracy and robustness of our approach.

1 Introduction

Most of the interventions of vessel diseases are performed by using catheters or guide-wires. One of the main difficulties is the navigation of these wires through the vessel system to the affected vessel. The common way of doing this is using a contrast agent and subtracted 2d X-ray images from a multi-axis angiography system (Fig. 2, left). Obviously, navigating a wire in a complex vessel system (e. g. brain, Fig. 2, middle left) using 2d X-ray images is not an easy task. Very good anatomic knowledge and a proper spatial sense are required, since the X-ray images do not contain any depth information. A 3d reconstruction of the current wire position inside of a 3d patient data set would be a great support for this difficult navigation task. Uncertainties during the navigation could be easily dissolved by rotating or zooming the 3d reconstruction.

We present a method for realtime online 3d reconstruction of a wire in a 3d patient data set (3d rotational angiography, 3DRA) using subtracted 2d X-ray images. The projection parameters are well known and the X-ray images do not have to be recorded from different viewpoints. This is very important since physicians do not want to change their current projection to get a 3d reconstruction of the guide-wire. They expect the reconstruction to be an additional feature of the single position fluoroscopy. Furthermore, we do not need special catheters or sensors. Given the reconstructed 3d patient vessel system, we use

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 386–393, 2009.

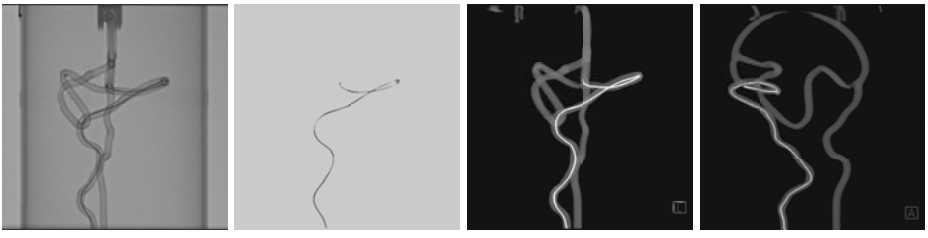


Fig. 1. Guide-wire in a vessel phantom (`phantom1`): native X-ray image (left), subtracted X-ray image (middle left) and two views of the 3d reconstruction of the estimated wire in a 3DRA (middle right, right)

recursive probability density propagation [1] to build a 3d probability distribution of the current 3d position of wire parts. The representation and temporal update of this distribution is performed by using a particle filter [2]. We extract the optimal 3d wire position using regularization techniques [3] based on this probability distribution. The wire is represented by uniform cubic B-splines [4]. Fig. 1 shows an X-ray image (left), the subtracted image (middle left) and the resulting reconstruction of the estimated 3d wire position inside of a 3DRA (middle right, right).

In the literature, one can find different approaches for the reconstruction of wires in a 3DRA. Solomon et al. [5] and Woods et al. [6] use special catheters and sensors to determine and visualize the 3d position of a catheter tip. Bender et al. [7] use two or more X-ray images from different viewpoints to build a 3d reconstruction of the guide-wire. By manually providing a starting point and direction of the wire, they reconstruct it iteratively using backprojection and comparing the backprojection to gradient images of the X-ray images. Baert et al. [8] use a calibrated biplane angiography system for 3d reconstruction of the guide-wire by triangulation. Therefore they need an accurate tracking of the guide-wire in the 2d X-ray images. Point correspondences for triangulation are extracted by using the epipolar constraint. By using an angiography system and a 3DRA van Walsum et al. [9] estimate the 3d guide-wire position by reprojecting the segmented wire from the 2d X-ray image. In opposite to our work, they do not use any temporal information for the 3d position estimation. We also do not need any special hardware, manual interaction, or an accurate 2d segmentation of the wire.

The remainder of this paper is structured as follows: In section 2 we describe our approach for estimating the 3d position of a wire using 2d X-ray images. We present different experiments on simulated and phantom data in section 3. Conclusions are given in section 4.

2 Method

The method we propose consists of two parts. First, we build with every new X-ray image a probability distribution of the currently possible 3d positions of

wire parts using the 3DRA, the prior knowledge of the wire position, i. e. the probability distribution of the last timestep, and the prior 2d X-ray images. This is done by recursive probability density propagation using a particle filter [2]. In combination with a state estimator (e. g. maximum a posteriori or maximum likelihood), a particle filter is a powerful tool for state estimation. Platzer et al. [10] use a particle filter for 3d blood flow reconstruction from 2d angiograms. Unfortunately, using a particle filter for state estimation is – due to practical reasons – only possible if the state space is low dimensional. Since a generic, complex 3d path of a wire cannot be described by just a few parameters, it is not promising to apply regular state estimation techniques. However, a particle filter still provides an efficient way to describe and update multi-modal probability distributions. The second part of our method consist of extracting the optimal wire position from this probability distribution using regularization techniques.

2.1 Probabilistic Estimation of the 3d Wire Position

Projecting 3d data into a 2d image plane causes information loss and creates ambiguities. Recursive probability density propagation [1] allows dissolving such ambiguities by integrating temporal information. In the context of our problem, it can be written as

$$p\left(\boldsymbol{x}_t \mid \langle \boldsymbol{o} \rangle_t\right) \propto \underbrace{p\left(\boldsymbol{o}_t \mid \boldsymbol{x}_t\right)}_{\text{likelihood}} \int \underbrace{p\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \langle \boldsymbol{o} \rangle_{t-1}\right)}_{\text{update}} \underbrace{p\left(\boldsymbol{x}_{t-1} \mid \langle \boldsymbol{o} \rangle_{t-1}\right)}_{\text{recursion}} \, \mathrm{d}\boldsymbol{x}_{t-1}. \quad (1)$$

Where $\boldsymbol{x}_t \in \mathbb{R}^3$ is some point in the 3DRA and $\langle \boldsymbol{o} \rangle_t = \langle \boldsymbol{o}_1, \boldsymbol{o}_2, \ldots, \boldsymbol{o}_t \rangle$ are the observations, i. e. subtracted X-ray images of the wire, until some point in time t. Recursive probability density propagation consists of three parts. The *recursion* $p(\boldsymbol{x}_{t-1} \mid \langle \boldsymbol{o} \rangle_{t-1})$ is the reason why only the current observation is needed to update the distribution, i. e. the knowledge of the last timestep. Due to a lack of information we choose the initial 3d wire distribution $p(\boldsymbol{x}_0)$ to be a uniform distribution inside of the patient vessel system.

In order to predict, where wire parts could move to in the next time step, the *update* probability distribution

$$p\left(\boldsymbol{x}_t \mid \boldsymbol{x}_{t-1}, \langle \boldsymbol{o} \rangle_{t-1}\right) \propto \frac{\mathcal{N}\left(d(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}) \mid \mu, \sigma^2\right)}{\int\limits_{\boldsymbol{\epsilon} \in \Omega} \int \mathcal{N}\left(d(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}) \mid \mu, \sigma^2\right) p\left(\boldsymbol{x}_{t-1} \mid \langle \boldsymbol{o} \rangle_{t-1}\right) \, \mathrm{d}\boldsymbol{x}_{t-1} \, \mathrm{d}\boldsymbol{\epsilon}}, \quad (2)$$

is used. In the literature [1], the update distribution does not depend on the old observations $\langle \boldsymbol{o} \rangle_{t-1}$. However, eq. (1) can be easily transformed to integrate the old observations into the update distribution. The normal distribution in the numerator and denominator with mean μ and variance σ^2 describes where wire parts could move to, given their current position. Important for this distribution is the shortest distance $d(\boldsymbol{x}_t, \boldsymbol{x}_{t-1})$ between the past and the current point considering the vessel structure. This shortest distance calculation is based on the 3DRA using graph-theoretical methods. Since we suppose that the wire never

leaves the vessel system, the distance to some point $\boldsymbol{x}_t$ outside of it is infinite. The denominator of eq. (2) prevents single high peaks in the 3d wire probability. This is achieved by decreasing the probability that wire parts move to positions where the 3d wire probability is already high. This normalization is restricted to a local space Ω around $\boldsymbol{x}_t$. The size of this space affects the smoothing. Note that this is very important, since otherwise single parts of the wire would get a much higher probability than other parts, which conflicts the prior information that each part of the wire should have the same probability. For the realisation using a particle filter, this means that it might happen that the particles tend to cluster at a small space instead of along the complete wire. This is because with every iteration more particles are drawn from the area with the higher probability.

The third part of the recursive probability density propagation is the *likelihood*

$$p(\boldsymbol{o}_t|\boldsymbol{x}_t) \propto \max_{(i^*,j^*)} \boldsymbol{o}_t\,(i^*,j^*)\,\mathcal{N}\,(i-i^*, j-j^* \mid \mathbf{0}, \boldsymbol{\Sigma}) \tag{3}$$

which is used to integrate the actual observation $\boldsymbol{o}_t$, i. e. the subtracted X-ray image, into the 3d wire probability distribution. Where (i, j) are the image coordinates of the projected 3d point $\boldsymbol{x}_t$ and $\boldsymbol{o}_t\,(i, j)$ is the gray value of the observation at this position weighted by a Gaussian kernel $\mathcal{N}\,(\cdot, \cdot \mid \mathbf{0}, \boldsymbol{\Sigma})$. The projection parameters are known by the angiography system which recorded the 2d X-ray image. In order to prevent higher probabilities in areas where the projected wire is crossing itself, we use the maximum instead of a sum.

2.2 Extraction of the Optimal Wire

We use uniform cubic 3d B-splines [4] $\mathbf{S}_\mathcal{C}(r) : [0, 1] \to \mathbb{R}^3$ to describe the 3d wire position. The appearance of a B-spline is defined by the position of its control points $\mathcal{C} = \{\boldsymbol{c}_0, \boldsymbol{c}_1, \ldots, \boldsymbol{c}_m\}$, $\boldsymbol{c}_k \in \mathbb{R}^3$. As start and end of the spline, we use the first and last control point, respectively: $\mathbf{S}_\mathcal{C}(0) = \boldsymbol{c}_0$, $\mathbf{S}_\mathcal{C}(1) = \boldsymbol{c}_m$.

Given the probability distribution $p(\boldsymbol{x}_t \mid \langle \boldsymbol{o} \rangle_t)$, which is caused by the wire, we want to extract the 3d wire position, i. e. the spline control points. Obviously this is an ill-conditioned problem which cannot be solved without specifying further constraints. But it can be solved using regularization techniques [3]. We search for the optimal spline control points

$$\hat{\mathcal{C}} = \operatorname*{argmax}_{\mathcal{C}} \underbrace{\int_0^1 \log p\left(\mathbf{S}_\mathcal{C}\,(r) \mid \langle \mathbf{o}_t \rangle\right)\,\mathrm{d}r}_{\text{probability}} + \alpha \underbrace{\int_0^1 \left|\mathbf{S}'_\mathcal{C}\,(r)\right|\,\mathrm{d}r}_{\text{length}} - \beta \underbrace{\int_0^1 \mathbf{S}''_\mathcal{C}\,(r)^2\,\mathrm{d}r}_{\text{smoothness}} \tag{4}$$

by maximizing an objective function which consists of three parts. The searched spline should stride an area with high *probability* and maximize *length* and *smoothness*. The parameters α and β weight the different parts of the objective function. Since the probability of any 3d point outside of the vessel system is zero, the spline cannot leave it. We use Powell's method [11] for optimization.

3 Experiments

We tested our method on two simulated sequences based on the 3DRA of a half brain (`sim1` and `sim2`, 300 images each) and on two sequences using real X-ray images (Fig. 1, left) of a vessel phantom (`phantom1`, 186 images and `phantom2`, 231 images). Fig. 2 (middle left, middle right) shows the two 3d data sets. Each of these data sets is scaled to a size of 256^3 voxels resulting in a voxel size of 0.765 mm for the phantom and 0.539 mm for the brain data set. An impression of the different sequences is given in Fig. 1 and Fig. 2 (right). Each of these images shows the estimated 3d wire position after the last image from the viewpoint where the 2d X-ray images of the respective sequence are taken. The 3d data sets and the X-ray images are recorded on a Siemens Artis zee C-arm system (Fig. 2 (left), image resolution: 1240×960, pixel size $0.308\,\text{mm} \times 0.308\,\text{mm}$). We use a 33×33 Gaussian kernel to build the 2d probability distribution of eq. (3). We choose $\mu = 7.5$ and $\sigma^2 = 2.5$ for eq. (2) and our particle set consists of 15000 particles. To reach realtime performance in our experiments, we use a graph, which is based on the thinned vessel system of the 3DRA, for the distance calculation inside of the vessel system. This graph is also used to find the longest possible wire given the 3d wire probability distribution what simplifies the extraction of the optimal wire and to decide how many spline control points are necessary for the wire representation.

In order to show the accuracy of our method, we analyze the error of the wire tip and the complete wire. The ground truth 3d and reprojected 2d wire position of the simulated sequences are well known. For the phantom sequences only the manually extrated 2d ground truth is given. Thus we analyze just the accuracy of the reprojected wire of these sequences. We calculate the 2d and 3d Euclidean distance between the estimated wire tip and the ground truth wire tip. To evaluate the error of the complete wire, we calculate the shortest Euclidean distance between each pixel (voxel) of the estimated wire to some pixel (voxel) of the ground truth wire. Note that this evaluation rates an estimated wire which is too short in comparison to the ground truth wire too good. However, this case

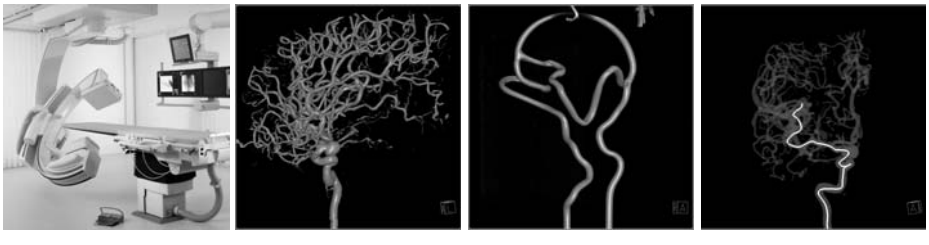


Fig. 2. Siemens Artis zee C-arm system (left, www.siemens.com). 3d half brain data set used for the simulated (middle left) and the 3d phantom data set used for the phantom sequences (middle right). 3d visualization of the estimated wire after the last image of the sequence `sim1` of subtracted X-ray images (right). The viewpoint of this image is similar to the viewpoint where the X-ray images are taken.

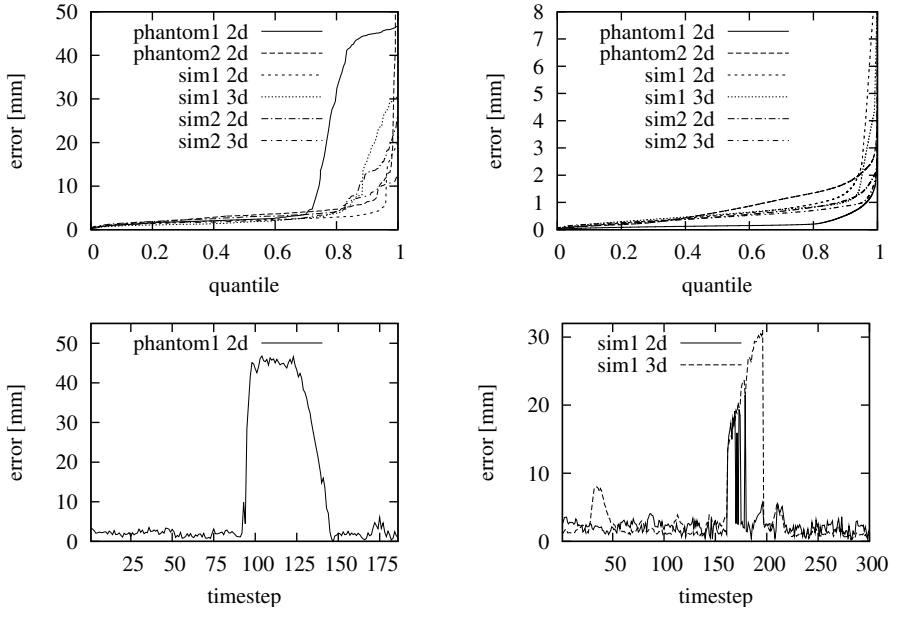


Fig. 3. Quantile plots of the 2d and 3d (only simulated sequences) error of the wire tip (top left) and the complete wire (top right) of each sequence. Error of the wire tip for each timestep of the sequences `phantom1` (bottom left) and `sim1` (bottom right). The overall median (quantile 0.5) error for the wire tip is 2.52 mm and 0.47 mm for the complete wire. Explanations for temporarily larger errors are given in section 3.1.

is handled by the wire tip error. Since the metric sizes of each pixel and voxel are known, we present the distances in millimeters.

Fig. 3 shows quantile plots of the wire tip error (top left) and of the complete wire error of each sequence (top right). An overview of the median (quantile 0.5) errors of each single sequence and of all sequences (`all`) is presented in Table 1. The median error of the reprojected wire tip is 2.52 mm and 0.47 mm for the complete wire. In section 3.1, we describe reasons for temporarily larger errors. Our approach takes 1.48 seconds for processing a timestep on an Intel Pentium 4 2.8 GHz with 1 GB RAM. Note that since a particle filter can be completely parallelized, much better runtimes are possible on a modern multi core system using an optimized implementation.

3.1 Reasons for Inaccuracies

A reason for small errors are ghosting artefacts. These artefacts are created by the X-ray detector due to a temporal averaging of the X-ray images. An example for a ghosting artefact is displayed in Fig. 4 (left). Because the wire is moved quite fast it seems to have two tips. Also most of the time the estimated wire tip position is slightly in front of the real wire tip position, since eq. (4) benefits long wires and the observation is slightly blurred.

Table 1. Median errors in millimeters for each single and for all sequences (`all`)

error [mm]	`phantom1`	`phantom2`	`sim1`	`sim2`	`all`
cath. tip 2d (3d)	2.27	3.24	2.22 (1.71)	2.90 (1.98)	2.52 (1.88)
catheter 2d (3d)	0.14	0.64	0.52 (0.55)	0.47 (0.45)	0.47 (0.50)

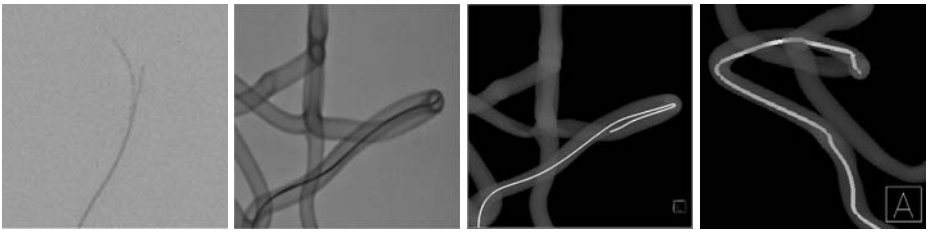


Fig. 4. Ghosting artefact (left), example for an ambiguity in `phantom1`: X-ray image (middle left), (incorrectly) estimated wire (middle right, right)

Most of the larger errors can be explained by ambiguities of the 3d catheter position in the 2d X-ray images. Fig. 3 (bottom left) shows the 2d error of the wire tip for each timestep of the `phantom1` sequence. Most of the time, this error is really low, but in the second part of the sequence, the error gets suddenly very large. The reason for this is the ambiguity shown in Fig. 4 (middle left). Because of an overlapping vessel, the 3d position of the wire cannot be estimated correctly. The (incorrectly) estimated wire is shown in Fig. 4 (middle right, right). Fig. 3 (bottom right) shows the 2d and 3d error of the wire tip in sequence `sim1`. At the start and in the middle of this sequence, the 3d error suddenly increases, while the 2d error stays low. The reason for this is that the wire is moving in the direction of the optical axis of the X-ray detector. Hence the 3d position of the wire tip cannot be estimated correctly. However, the reprojected 2d position is still quite accurate.

Note that in each of these cases our method is robust enough to return very quickly to a high accuracy as soon as the ambiguities are resolved. Furthermore, many of these ambiguities can easily be resolved by using a second pair of X-ray source and detector with a different viewpoint, which could be easily provided by a biplane system. In this case, the likelihood in eq. (11) would consist of two observations.

4 Conclusions

We presented a new approach for realtime online 3d guide-wire reconstruction using 2d X-ray images, which do not have to be recorded from different viewpoints. This meets the clinical needs since physicians do not want to change their current projection to get a 3d reconstruction of the guide-wire. By using recursive probability density propagation, we built a 3d probability distribution of the

positions of wire parts based on a 3d patient data set and 2d X-ray images. In order to represent and update this multi-modal distribution, we used a particle filter. To extract the optimal 3d wire position from this distribution, we applied regularization techniques. We described the 3d wire position by a uniform cubic B-spline. Our method does not need any special catheters or sensors. Performing different experiments on simulated and phantom data showed the high accuracy of our method (Table 1). We achieved a median error of the reprojected wire tip of 2.52 mm. The median error of the reprojected complete wire was 0.47 mm. The achieved median 3d error is 1.88 mm for the tip and 0.50 mm for the complete wire.

In our future work, we will improve the accuracy of our approach by using a biplane angiography system. In order to better model the guide-wire behaviour we will also add a speed estimation of the guide-wire tipp. We will also implement our approach using multi processing techniques. The clinical evaluation of our approach is imminent.

References

1. Doucet, A., DeFreitas, N., Gordon, N.: Sequential Monte Carlo Methods in Practice. Springer, Heidelberg (2001)
2. Arulampalam, M., Maskell, S., Gordon, N., Clapp, T., Sci, D., Organ, T., Adelaide, S.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. on Signal Processing 50(2), 174–188 (2002)
3. Engl, H., Hanke, M., Neubauer, A.: Regularization of Inverse Problems. Kluwer Academic Publishers, Dordrecht (1996)
4. de Boor, C.: A Practical Guide to Splines. Springer, Heidelberg (1978)
5. Solomon, S., Dickfeld, T., Calkins, H.: Real-Time Cardiac Catheter Navigation on Three-Dimensional CT Images. J. of Interv. Card. Electrophys. 8(1), 27–36 (2003)
6. Wood, B., Zhang, H., Durrani, A., Glossop, N., Ranjan, S., Lindisch, D., Levy, E., Banovac, F., Borgert, J., Krueger, S., et al.: Navigation with Electromagnetic Tracking for Interventional Radiology Procedures: A Feasibility Study. J. of Vascular and Interv. Radiology 16(4), 493–505 (2005)
7. Bender, H.J., Männer, R., Poliwoda, C., Roth, S., Walz, M.: Reconstruction of 3D Catheter Paths from 2D X-ray Projections. In: Proc. of the Second Int. Conference on Med. Image Computing and Computer-Assisted Intervention, pp. 981–989 (1999)
8. Baert, S., van de Kraats, E., van Walsum, T., Viergever, M., Niessen, W.: Three-dimensional guide-wire reconstruction from biplane image sequences for integrated display in 3-D vasculature. IEEE Trans. on Med. Imaging 22(10), 1252–1258 (2003)
9. van Walsum, T., Baert, S., Niessen, W.: Guide wire reconstruction and visualization in 3DRA using monoplane fluoroscopic imaging. IEEE Trans. on Med. Imaging 24(5), 612–623 (2005)
10. Platzer, E.S., Deinzer, F., Paulus, D., Denzler, J.: 3D Blood Flow Reconstruction from 2D Angiograms. In: Bildverarbeitung für die Medizin 2008 - Algorithmen, Systeme, Anwendungen. Informatik aktuell, pp. 288–292. Springer, Heidelberg (2008)
11. Powell, M.J.D.: An efficient method for finding the minimum of a function of several variables without calculating derivatives. The Comp. J. 7(2), 155–162 (1964)

3-D Respiratory Motion Compensation during EP Procedures by Image-Based 3-D Lasso Catheter Model Generation and Tracking

Alexander Brost[1], Rui Liao[2], Joachim Hornegger[1], and Norbert Strobel[3]

[1] Chair of Pattern Recognition, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, Erlangen, Germany
Alexander.Brost@informatik.uni-erlangen.de
[2] Siemens Corporate Research, Imaging and Visualization, Princeton, NJ, USA
[3] Siemens AG, Healthcare Sector, Forchheim, Germany

Abstract. Radio-frequency catheter ablation of the pulmonary veins attached to the left atrium is usually carried out under fluoroscopy guidance. Two-dimensional X-ray navigation may involve overlay images derived from a static pre-operative 3-D volumetric data set to add anatomical details. However, respiratory motion may impair the utility of static overlay images for catheter navigation. We developed a system for image-based 3-D motion estimation and compensation as a solution to this problem for which no previous solution is yet known. It is based on 3-D catheter tracking involving 2-D/3-D registration. A biplane X-ray C-arm system is used to image a special circumferential (lasso) catheter from two directions. In the first step of the method, a 3-D model of the device is reconstructed. 3-D respiratory motion at the site of ablation is then estimated by tracking the reconstructed model in 3-D from biplane fluoroscopy. In our experiments, the circumferential catheter was tracked in 231 biplane fluoro frames (462 monoplane fluoro frames) with an average 2-D tracking error of 1.0 mm ± 0.5 mm.

1 Introduction

Atrial fibrillation (AF) is one of the most common heart rhythm disorders and a leading cause of stroke. Radio-Frequency catheter ablation (RFCA) has become an accepted option for treating AF in today's electrophysiology (EP) labs, in particular, if drug treatment has become ineffective. RFCA of the pulmonary veins (PVs) is usually carried out under fluoroscopy guidance. Unfortunately, X-ray projection images cannot distinguish soft tissue well. To address this issue, fused visualization combining pre-operative high-resolution 3-D atrial CT and/or MR volumes with the fluoroscopic images has been developed (fluoro overlay image guidance) [1]. Current fluoro overlay techniques are however usually static, i.e., they do not follow the heart while it beats and moves through the breathing cycle. This effect may impair the utility of fluoro overlays for catheter navigation. By synchronizing the fluoroscopic images with electrocardiogram (ECG), cardiac motion can be eliminated, and respiratory motion can be isolated for

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 394–401, 2009.

the fused visualization. While it has been widely recognized that motion compensation is crucial for fluoro overlays, image-based 3-D motion-compensation methods designed for EP applications are not yet known, mainly due to the fact that there are few discernible features in typical EP fluoro images. Motion compensated navigation for coronary intervention using magnetic tracking was suggested in [2], but it requires special catheters equipped with an electromagnetic sensor at increased cost. Only vertical motion in the imaging plane was compensated in [3] and [4] for liver embolization and hepatic artery catheterization, respectively. But the left atrium undergoes a three dimensional motion during respiration, as shown in [5]. This paper describes a novel image-based method to detect and compensate respiratory motion in 3-D using a biplane C-arm fluoroscopy device. Such an X-ray system has two imaging planes commonly referred to as A-plane and B-plane, respectively. To perform motion estimation, we track a commonly used EP catheter, the ring-shaped spiral catheter. This catheter is of unique shape and represents a good feature for tracking and motion estimation, as it is typically fixed at the ostium of the PV that is to be ablated.

2 Three-Dimensional Lasso Model Initialization

Our algorithm is based on the assumption that the perspective projection of a lasso catheter, when fit to the pulmonary veins, can be approximated as a 2-D ellipse. In turn, the lasso catheter is reconstructed in 3-D from biplane X-ray projections that are taken simultaneously using a calibrated biplane C-arm X-ray system. The lasso catheter itself can be approximated as an ellipse in 3-D, because a 3-D elliptical object remains elliptical when perspectively projected onto a 2-D imaging plane. Under some special view orientations, the ellipse can, however, collapse to a line. As a consequence, our algorithm is designed to reconstruct a 3-D lasso model either from 2-D ellipses extracted from biplane X-ray views, or from one ellipse and one line (Fig. 1).

2.1 Two-Dimensional Ellipse Extraction

The projection of the lasso catheter on the imaging plane is first extracted by manual clicking followed by fast marching in one frame of the fluoroscopy sequence, as explained in [6]. A 2-D ellipse is then fitted using the method described

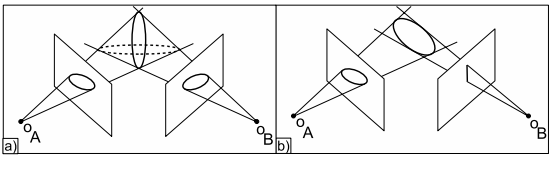


Fig. 1. 3-D lasso model initialization from two views. **a)** The 3-D model is found by the intersection of two 3-D elliptical cones. **b)** The model is found by intersecting the 3-D elliptical cone with the plane determined by the 2-D line and the optical center.

in [7]. Specifically, the method uses a least square estimation of the ellipse parameters $\mathbf{f} = (a, b, c, d, e, f)^T$ of the ellipse equation $au^2 + buv + cv^2 + du + ev + f = 0$ with the 2-D coordinates u and v. Given a set of points $\mathbf{p}_i = (u_i, v_i)^T$ with $i = 1, \ldots, N$, these can be combined in a measurement matrix $\mathbf{M}$ with the i-th row $\mathbf{M}(i,:) = (u_i^2, u_i v_i, v_i^2, u_i, v_i, 1)$ [8]. As the points may not necessarily lie exactly on the ellipse to be fitted, we are looking for the ellipse parameters $\hat{\mathbf{f}}$ that minimize $\hat{\mathbf{f}} = \arg\min_{\mathbf{f}} ||\mathbf{M}\mathbf{f}||_2^2$ subject to $||\mathbf{f}||_2 = 1$. Since the constraint $||\mathbf{f}||_2 = 1$ does not guarantee an elliptic solution, the method presented in [7] is used to assure the correct solution by enforcing the condition $b^2 - 4ac < 0$.

2.2 Three-Dimensional Ellipse Reconstruction

A 3-D point is projected onto an imaging plane as a 2-D point $\mathbf{p} = (u, v)^T$ by $\widetilde{\mathbf{p}} = \mathbf{P} \cdot \mathbf{w}$. This equation holds in homogeneous coordinates where $\mathbf{w} = (x, y, z, 1)^T$ and $\widetilde{\mathbf{p}} = (uw, vw, w)^T$. The projection matrix $\mathbf{P} \in \mathbb{R}^{3\times4}$ is defined by $\mathbf{P} = \mathbf{K}[\mathbf{R}|\mathbf{t}]$ with the matrix $\mathbf{K} \in \mathbb{R}^{3\times3}$ representing the intrinsic camera parameters, and the matrix $\mathbf{R} \in \mathbb{R}^{3\times3}$ and the vector $\mathbf{t} \in \mathbb{R}^3$ comprising the extrinsic camera parameters for rotation and translation, respectively. A 3-D elliptical cone can be spanned considering an imaging plane with the projection matrix P and the ellipse within that plane parameterized by $\mathbf{C} \in \mathbb{R}^{3\times3}$. The base of the elliptical cone is the ellipse in the imaging plane and the vertex is the optical center that can be calculated as $\mathbf{o} = -\mathbf{R}^{-1}\mathbf{t}$. It can be shown that the elliptical cone can be represented by $\mathbf{Q} = \mathbf{P}^T\mathbf{C}\mathbf{P}$ in matrix presentation [9]. For the regular case, the 3-D ellipse representing the 3-D lasso catheter is reconstructed by intersecting the two elliptical cones $\mathbf{Q}_A$ and $\mathbf{Q}_B$ corresponding to plane A and plane B of a biplane system respectively. The solution is found by calculating η so that the quadric $\mathbf{Q}(\eta) = \mathbf{Q}_A + \eta\mathbf{Q}_B$ is of rank 2 [9]. From a mathematical point of view, there are two feasible solutions, as illustrated in Fig. 1 a). In our application, we utilize our prior knowledge about the pseudo-circular shape of the lasso catheter and select the result that is more circular. For the special case where the lasso catheter is projected as close to being a line in one view, the method in [9] is not numerically stable. The special case is detected by considering the rank of the measurement matrix containing the first and second order point coordinates. A rank deficiency (ill-conditioning) indicates a linear dependency and therefore an ellipse that collapsed to a line. In this case, we propose to reconstruct the 3-D lasso model by the following procedures. First, a line is fitted to the line-like projection of the lasso catheter. It is calculated as the the principal axis of the points obtained by fast marching. Two arbitrary but distinct points $\mathbf{q}_1$ and $\mathbf{q}_2$ are then randomly selected on the fitted line and are connected to the optical center. The projection plane in which both the X-ray source and the fitted line lies is then determined by the two rays $\mathbf{q}_{1,o}$ and $\mathbf{q}_{2,o}$. In the second and final step, the 3-D lasso model is obtained by intersecting this plane with the elliptical cone defined by the ellipse from the second view. We propose the following method to calculate the analytical formulation of the conic section that intersects a plane $\widetilde{\mathbf{n}} = (x_n, y_n, z_n, d_n)^T$ with an elliptical cone $\mathbf{Q}$. The key idea is to transform the original coordinates, taken with respect to the iso-center of the C-arm system, to

new coordinates, in which the transformed x-y-plane coincides with the cutting plane $\widetilde{\mathbf{n}}$. The transformation is given as

$$\mathbf{S} = \begin{pmatrix} \mathbf{u}\ \mathbf{v}\ \mathbf{n}\ 0 \\ 0\ 0\ d_n\ 1 \end{pmatrix} \tag{1}$$

where $\mathbf{n} = (x_n, y_n, z_n)^T$ is normalized to unit length, and $\mathbf{u}$ and $\mathbf{v}$ can be any pair of unit length vectors lying in the cutting plane that are orthogonal to each other. $\mathbf{S}$ therefore presents a standard rigid-body transformation matrix after which the conic section $\mathbf{w}' = \mathbf{S}\mathbf{w}$ lies in the transformed x-y-plane with $z' = 0$. In addition, since

$$(\mathbf{w}')^T \mathbf{U} \mathbf{w}' = 0 \qquad \text{with} \qquad \mathbf{U} = ((\mathbf{S}^{-1})^T \mathbf{Q} \mathbf{S}^{-1}) \tag{2}$$

and considering $z' = 0$, the parameters of the analytical formulation of the conic section can be obtained straightforwardly from the matrix $\mathbf{U}$ by $\hat{a} = u_{1,1}$, $\hat{b} = 2u_{1,2}$, $\hat{c} = u_{2,2}$, $\hat{d} = 2u_{1,4}$, $\hat{e} = 2u_{2,4}$, $\hat{f} = u_{4,4}$ where $u_{i,j}$ represents the element of the i-th row and j-th column of matrix $\mathbf{U}$. The model points in the transformed coordinate system are given as $\mathbf{w}'_i$ (in homogeneous coordinates), $i = 1, \ldots, L$, with the number of model points L, and are calculated using the estimated ellipse parameters $\hat{a}, \hat{b}, \hat{c}, \hat{d}, \hat{e}, \hat{f}$ as explained above. The 3-D points for the model of the lasso catheter in the C-arm isocenter coordinate system, $\mathbf{w}_i$ (in homogeneous coordinates), are then calculated by $\mathbf{w}_i = \mathbf{S}^{-1}\mathbf{w}'_i$.

3 Three-Dimensional Lasso Model Tracking

After the 3-D lasso model has been generated from the first frame of the fluoro sequence, it is tracked in 3-D throughout the remainder of the biplane sequence. To speed up the computational efficiency and to minimize the influence of peripheral structures that could interference with lasso tracking, the region of interest (ROI) for tracking is restricted to 400 X 400 pixels (on the 1024 X 1024 image) around the center of the tracked lasso catheter in the previous frame. Histogram equalization is further applied on the ROI to enhance image contrast. In addition, a vesselness filter [10] is used to enhance line-like structures such as the lasso catheter. The feature image is then binarized using Otsu's algorithm [11]. Finally, a distance map is calculated from the binarized image [12]. A distance map encodes the distance from a point to its closed feature point, that is the nonzero point representing the extracted lasso catheter in our binarized feature image. The distance transform offers an important advantage. It provides a denoised representation of the fluoro image with a pronounced minimum around the 2-D shape of the lasso catheter. This way, we can still reach a good registration, even if we have a small 3-D model error, or if we start from a position that is somewhat away from the lasso catheter to be tracked. Lasso model tracking in 3-D is achieved by performing 2-D/3-D registration. To this end, the reconstructed 3-D lasso model is rotated by R and translated by T in 3-D first. Then it is projected onto the two imaging planes of the biplane C-arm system. The

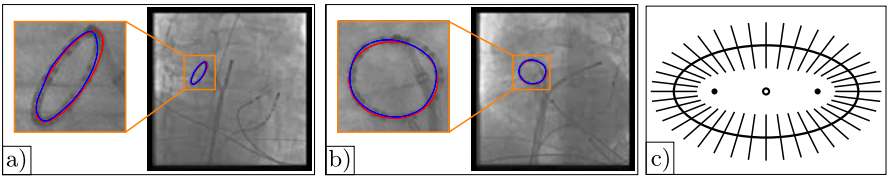


Fig. 2. a) Tracked ellipse (red) in plane A has an average distance to the manual segmentation (blue) of 1.0 mm (model error 0.5 mm). **b)** In this particular frame, the tracking error (0.6 mm) is all due to the model error (0.6 mm). **c)** Ellipse with center (circle), focal points (dots) and normals.

average distance between the projected points and the closest feature point (i.e. the lasso catheter) in fluoroscopic images is efficiently calculated using the distance map introduced above. A suitable rotation and translation is found by optimizing

$$\hat{\mathbf{R}}, \hat{\mathbf{T}} = \arg\min_{\mathbf{R},\mathbf{T}} \sum_i \mathbf{I}_{\mathrm{DT}}(\mathbf{P}_A \cdot \mathbf{T} \cdot \mathbf{R} \cdot \mathbf{w}_i) + \sum_i \mathbf{I}_{\mathrm{DT}}(\mathbf{P}_B \cdot \mathbf{T} \cdot \mathbf{R} \cdot \mathbf{w}_i). \quad (3)$$

However, due to the fact that the shape of lasso catheters may not always be exactly elliptical, a simple elliptical 3-D model may not fit perfectly. To still obtain robust tracking, the distance of a forward projected 3-D point to the closest feature point is calculated as the smallest distance among all the points along the normal direction within five pixels from the projected point. An illustration of the normals to an ellipse is given in Fig. 2 c). A best neighbor optimizer is used to iteratively optimize the translational and rotational parameters. Registration is performed in two steps. In the first step, only the translation is considered, whereas in the second step a fully six-dimensional optimization is performed. Two-step registration is implemented to increase performance by lowering the number of iterations required for optimization.

4 Results

We evaluated our algorithm by calculating the tracking error for each of the imaging planes throughout seven different clinical fluoro sequences that were acquired during EP procedures on an AXIOM Artis dBA C-arm system (Siemens AG, Healthcare Sector, Forchheim, Germany). The system was calibrated using the method presented in [13]. We focus on a typical setup involving one lasso catheter and one ablation catheter. The presence of other structures should not decrease the accuracy of our method, because we use a unique elliptical structure for registration. Model generation was experimentally evaluated by adding Gaussian noise with a standard deviation of 5 mm to 2-D forward-projected (3-D) ellipse points. In this case, we found the average distance between the resulting 3-D model and the initial 3-D model to be 3.2 mm. To calculate the tracking

Table 1. Average tracking error for the clinical sequences used. The last row shows an average over all the 7 sequences for plane A and plane B, respectively.

No.	Plane A			Plane B		
	Mean	Std.	Model Error	Mean	Std.	Model Error
1	1.0 mm	0.3 mm	0.8 mm	1.0 mm	0.4 mm	0.5 mm
2	1.1 mm	0.3 mm	0.5 mm	0.7 mm	0.2 mm	0.6 mm
3	0.9 mm	0.3 mm	0.4 mm	0.9 mm	0.4 mm	0.3 mm
4	1.0 mm	0.2 mm	0.6 mm	1.8 mm	0.4 mm	1.1 mm
5	1.1 mm	0.2 mm	0.8 mm	2.1 mm	1.0 mm	0.7 mm
6	1.0 mm	0.6 mm	0.5 mm	0.8 mm	0.2 mm	0.6 mm
7	1.0 mm	0.2 mm	0.5 mm	1.1 mm	0.2 mm	0.7 mm
Σ	1.0 mm	0.5 mm	0.6 mm	1.0 mm	0.5 mm	0.6 mm

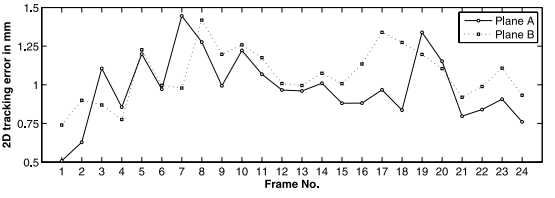


Fig. 3. Two-dimensional tracking error in mm for one sequence (No. 7) frame by frame. The A-plane and B-plane model errors are 0.5 mm and 0.7 mm, respectively.

error, we forward projected the 3-D catheter model, computed from first frame, into both planes of the biplane imaging system after 2-D/3-D registration had been performed. In each frame, we calculated the average distance of the forward projected 3-D catheter model to a manually segmented lasso catheter. The manual catheter segmentation was supervised by a cardiologist, and we consider it our reference. This distance of the forward projected 3-D model to the reference was averaged over all frames of a particular sequence to arrive at an overall 2-D tracking error for each sequence. Both A-plane and B-plane tracking results for each of the seven test sequences are summarized in Table 1. The mean average over the mean tracking errors was 1.0 mm $\pm$ 0.5 mm for plane A and 1.0 mm $\pm$ 0.5 mm for plane B. An example for one frame of one sequence is given in Figs. 2 a) and b). Our method currently achieves a frame rate of 3 frames-per-second using a (single threaded) CPU implementation. Since this frame rate is clinically used for EP procedures to keep X-ray dose low, real-time tracking can be achieved in such a situation. In addition to the tracking error, a model error can be calculated. The model error can be obtained from the first image of a sequence as no registration was performed in this particular frame. A detailed frame-by-frame 2-D tracking error for sequence no. 7 is given in Fig. 3.

5 Discussion and Conclusions

Our experiments on clinical EP fluoro sequences show that the tracking error averages 1 mm including an average model error of 0.6 mm. The real tracking

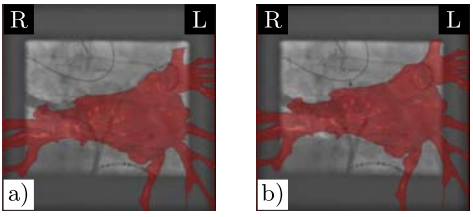


Fig. 4. a) Frame of a non-motion compensated sequence with fluoro overlay. **b)** The same sequence with motion compensation. Note how well both lasso catheter and contrast agent are matched to the morphology of the left atrium.

accuracy is therefore sub-millimeter. Considering that breathing motion in typical EP fluoro images is in the range of 15 mm and for some patients can be up to 40 mm for deep breathing, our method has the potential to significantly improve the accuracy of fluoro overlay images for EP navigation. To the best of our knowledge, this is the first image-based method that is specifically designed for the challenging task of 3-D breathing motion compensation during EP applications. The proposed method offers several advantages. First, it is workflow-friendly and does not require fiducial markers or additional contrast agent to be administered. Second, breathing motion is estimated directly in 3-D and right at the ablation site, meaning that motion estimation and compensation is essentially done in one step. There is no ambiguity coming from the inference of the real motion from surrogate motion estimates. Third, we estimate the motion online and update it constantly from fluoroscopy. Therefore we do not rely on a predefined motion model from which the real motion may deviate significantly during the procedure. Fourth, there are no restrictions for the 3-D data set that can be used as a fluoro overlay. In other words, the 3-D data set could come from MRI, CT or C-arm CT, e.g., *syngo* DynaCT Cardiac (Siemens AG, Healthcare Sector, Forchheim, Germany). Since the motion of the left atrium can be approximated by a rigid-body transform [5], it is possible to apply the motion estimate obtained by 3-D catheter tracking to the static fluoro overlay. This way, we can obtain an animated version of our initial overlay that moves in sync with the real anatomy. Fig. 4 a) represents the conventional overlay technique without motion compensation, while Fig. 4 b) shows an animated fluoro overlay with motion compensation. With motion compensation, 3-D fluoro overlay and lasso catheter, fixed at the PV ostium, remain in sync. In addition, the contrast enhanced upper pulmonary vein, shown on the left side of the fluoro image in Fig. 4 b), remains well aligned with the corresponding vein of the volumetric data. Since simultaneous biplane imaging for EP applications is usually performed with both imaging planes set orthogonal to each other, tracking in 3-D should be even more accurate than suggested by our 2-D error analysis.

References

1. Zagorchev, L., Manzke, R., Cury, R., Reddy, V.Y., Chan, R.C.: Rapid fusion of 2D x-ray fluoroscopy with 3D multislice CT for image-guided electrophysiology procedures. In: Medical Imaging 2007: Visualization and Image-Guided Procedures, San Diego, CA, USA, SPIE, vol. 6509, 65092B (2007)
2. Timinger, H., Krueger, S., Dietmayer, K., Borgert, J.: Motion Compensated Coronary Interventional Navigation by Means of Diaphram Tracking and Elastic Motion Models. Phys. Med. Biol. 50(3), 491–503 (2005)
3. Ross, J.C., Subramanian, N., Solomon, S.B.: Motion Correction for Augmented Fluoroscopy - Application to Liver Embolization. In: 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro. ISBI, pp. 1553–1556. IEEE Computer Society Press, Los Alamitos (2008)
4. Atasoy, S., Groher, M., Zikic, D., Glocker, B., Waggershauser, T., Pfister, M., Navab, N.: Real-time Respiratory Motion Tracking: Roadmap Correction for Hepatic Artery Catheterizations. In: Miga, M.I., Cleary, K.R. (eds.) Medical Imaging 2008: Visualization, Image-guided Procedures, and Modeling, San Diego, CA, USA, SPIE, vol. 6918, p. 691815 (2008)
5. Ector, J., Buck, S.D., Loeckx, D., Coudyzer, W., Maes, F., Dymarkowski, S., Bogaert, J., Heidbüchel, H.: Changes in Left Atrial Anatomy Due to Respiration: Impact on Three-Dimensional Image Integration During Atrial Fibrillation Ablation. Journal of Cardiovascular Electrophysiology 19(7), 828–834 (2008)
6. Liao, R., Xu, N., Sun, Y.: Location Constraint Based 2D-3D Registration of Fluoroscopic Images of CT Volumes for Image-Guided EP Procedures. In: Miga, M.I., Cleary, K.R. (eds.) Medical Imaging 2008: Visualization, Image-guided Procedures, and Modeling, SPIE, vol. 6918, p. 69182T (2008)
7. Halír, R., Flusser, J.: Numerically Stable Direct Least Squares Fitting Of Ellipses. In: Proceedings of the 6th Conference in Central Europe on Computer Graphics and Visualization, University of West Bohemia, Campus Bory, Plzen - Bory, Czech Republic, February 1998, pp. 253–257 (1998)
8. Hartley, R.I., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
9. Quan, L.: Conic Reconstruction and Correspondence From Two Views. IEEE Transactions On Pattern Analysis And Machine Intelligence 18(2), 151–160 (1996)
10. Sato, Y., Nakajima, S., Shiraga, N., Atsumi, H., Yoshida, S., Koller, T., Gerig, G., Kikinis, R.: 3D Multi-Scale Line Filter for Segmentation and Visualization of Curvilinear Structures in Medical Images. Medical Image Analysis 2(2), 143–168 (1998)
11. Otsu, N.: A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics 9(1), 62–66 (1979)
12. Breu, et al.: Linear time Euclidean distance transform algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(5), 529–533 (1995)
13. Rougee, A., Picard, C.L., Trousset, Y.L., Ponchut, C.: Geometrical calibration for 3D x-ray imaging. In: Kim, Y. (ed.) Medical Imaging 1993: Image Capture, Formatting, and Display, Newport Beach, CA, USA, SPIE, February 1993, vol. 1897, pp. 161–169 (1993)

System Design of a Hand-Held Mobile Robot for Craniotomy

Gavin Kane[1], Georg Eggers[1], Robert Boesecke[1], Jörg Raczkowsky[2], Heinz Wörn[2], Rüdiger Marmulla[1], and Joachim Mühling[1]

[1] Klinik und Poliklinik für Mund-, Kiefer, und Gesichtschirurgie, University of Heidelberg, Heidelberg, Germany
[2] Institute for Process Control and Robotics, University of Karlsruhe, Karlsruhe, Germany

Abstract. This contribution reports the development and initial testing of a Mobile Robot System for Surgical Craniotomy, the Craniostar. A kinematic system based on a unicycle robot is analysed to provide local positioning through two spiked wheels gripping directly onto a patients skull. A control system based on a shared control system between both the Surgeon and Robot is employed in a hand-held design that is tested initially on plastic phantom and swine skulls. Results indicate that the system has substantially lower risk than present robotically assisted craniotomies, and despite being a hand-held mobile robot, the Craniostar is still capable of sub-millimetre accuracy in tracking along a trajectory and thus achieving an accurate transfer of pre-surgical plan to the operating room procedure, without the large impact of current medical robots based on modified industrial robots.

1 Introduction

There are a number of surgical craniotomy procedures in which it is desired for a precise pre-surgical plan to be accurately transferred to the Operating Room (OR), such procedural examples include minimally invasive neuro-surgical procedures where the desired entry hole is pre-planned, frontal orbital advancement procedures for maxillo-facial surgery where the desired cranium advancement is pre-planned, or in a more recent example the milling of bone surfaces in plastic surgery according to a three dimensional operation plan [1] or CAD/CAM prefabricated skull implants requiring accurate bone resection for placement [2]. In support of this requirement for accurate transfer, many research groups have developed robotic solutions to address the specific challenges associated with the conduct of a craniotomy procedure [2], [3], [4]. To date, none of these solutions have been accepted for commercial use in performing craniotomys, it is assessed due to three facts:

- The surgical robotic systems presented are modified industrial robots whose impact within the OR in terms of footprint and the required changes to surgical workflows and procedures is quite considerable.
- The risk inherent with the demonstrated robotic craniotomys are considerably large, with respect to cutting too deep, causing meninges tears, and possibly thereafter damage to the brain.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 402–409, 2009.

- The robotic solutions offered are all supervisory controlled interventions where, by the definition of Nathoo et al. [5], the robot performs the pre-planned and programmed movements autonomously, thus removing the surgeon from the procedure. This occurring at the key time when his / her years of experience and 'feel' for the operation are most valuable.

Shared control is not a new concept in Surgical Robotics. Synergistic control was proposed by Taylor et al. [6] as early as 1991. Recent examples of this control in surgical robots include the Acrobot [7] and the PADyC system [8]. However, these two systems are still modified industrial robots, whose large footprint and individualized workflows still create a substantial impact to the OR. There are also no similar systems developed for the conduct of craniotomys.

There is an alternate trend in development, which is heading towards the minimisation of the impact of robotics in the OR: through the use of intelligent hand-held tools. Example projects in this area include the Precision Freehand Sculptor (PFS) by Brisson et al. [9] and the Sicherheits Trepenation System (STS) by Follman et al [10]. The PFS demonstrates how bone segments, such as those for knee implants, can be accurately removed according to a pre-planned model. On the other hand, the STS demonstrates how a normal hand-held drill can be enhanced for safety with the addition of both intrinsic and extrinsic system sensors. Unfortunately, this second system is limited to safety enhancement, and is not capable of any assisting the Surgeon in achieving an accurate pre-planned trajectory in line with the previously stated requirements.

From all the above examples, there is still no developed capability for a safe shared control system for craniotomys, specifically capable of accurately milling pre-planned trajectory in the OR with minimal impact on OR footprint and workflows. The solution presented here is a novel approach to Surgical Robotics, which addresses the following hypothesis:

1. A mobile robot, built around a standard craniotomy drill piece, hand-held by the surgeon, can drive the drill piece around a skull, on a pre-defined trajectory.
2. The system shall be capable of achieving surgical accuracy with the use of spiked wheels to ensure the skull remains a non-slip surface for dynamic analysis.
3. The control system supporting the robot is capable of guiding the surgeons hand during a procedure, but is not to remove him/her from complete control. i.e. This is to be a shared control system by the definition in [5].

In this paper we present the design, theory and performance testing of the prototype robot system, the 'Craniostar' shown in Fig. 1.

2 Methods and Materials

One of the concepts for using a mobile hand-held robot discussed above was the minimisation of the impact into the Operating Room. The components accepted as standard for Image Guided Surgery (IGS) include a tracking system and imaging support (GUI). All attempts were then made to minimise additional footprint requirements (that of the robot and robot control) but maximise the integration of the system into the existing OR. A key component of this integration was the generation of an

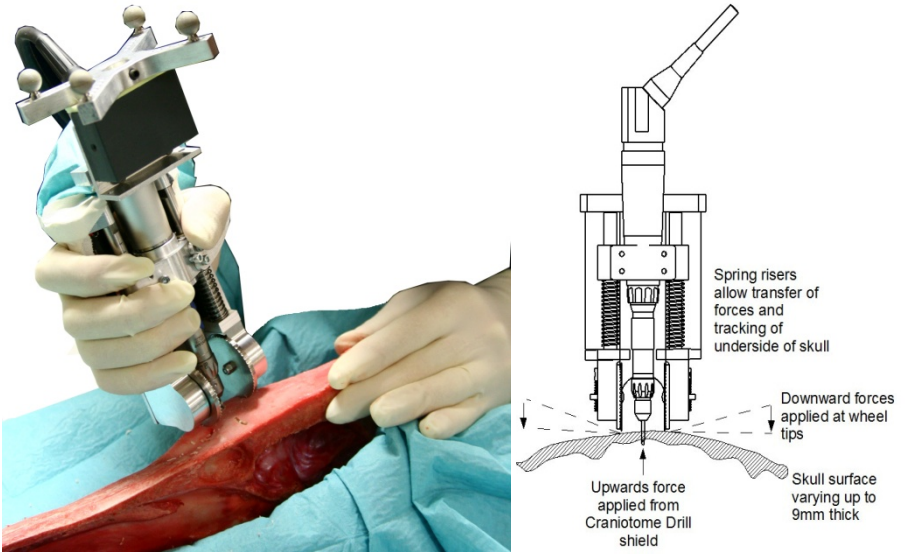


Fig. 1. (a) The prototype mobile hand-held surgical robot, 'Craniostar' being tested on a cadaver, and (b) a CAD drawing showing the principle forces used to ensure traction

intuitive Human / Computer Interface (HCI). The HCI is based on employing three inputs for the robot, from components already existing inside the OR. The first is the use of the Navigation System. The navigation system detecting minor tilts of the robot, occurring when the Surgeon pushes the robot forward implying an intent to move faster, or pulling the robot backwards implying an intent to slow down or stop. The second and third inputs are from the existing high speed drill, the foot pedal input for the desired speed of the drill, and the torque feedback of the drill for the difficulty the robot is facing in cutting. By integrating this control into the system, [11] showed that a minimal impact, intuitive shared control could be achieved.

The overall concept for employment follows a generalised IGS workflow with commencement of pre-surgical imagery and planning. Here a CT is employed to support the surgeons planning the Craniotomy through the KASOP software[12]. The location for the start of the Craniotomy is determined using the GUI and either an optically tracked pointer or the tip of the Craniostar itself, and the first burr hole is made. If required the Dura matter can be partially separated from the underside of the skull. If a circular craniotomy is desired, no further burr holes are required, and the tip of the Craniostar is inserted facing roughly in the correct direction. For a two ended craniotomy, a second bore hole is required at the finish location. The control of the movement of the Craniostar along the trajectory is discussed in the following sections on steering and control.

Firstly, this chapter will discuss how this system was achieved through a custom mechanical design, and then the control system is discussed.

2.1 Mechanical Design

The concept for the mobile robot, shown in Fig. 1, comprises a platform (robot body), two wheels, and the craniotomy high speed drill directly at the centre of the wheels' axis, with an optical tracking marker placed on the top of the robot. The high speed drill (Aesculap AG, Tuttlingen, Germany) is held in a brace in the centre of the platform. The brace rests on two spring supported sliders, allowing the highspeed drill to move up and down tracking the base of the skull. The spring force generated on the sliders has two purposes, firstly to prevent the drill and Dura Guard being pushed hard against the Dura, and secondly in order to transfer force to the wheels of the robot, supporting the friction of the wheels and hence the traction. The required spring strength was calculated based on the ability of the Dura Guard to track the bottom of the skull, with depth changes to a maximum of 9mm while moving up to a maximum cutting speed of 5mm/s. The Dura Guard being a small metal hook that extends underneath the drill piece, preventing the drill from cutting into the Dura.

The motor drive is achieved through two 25W Motors (Maxon Motors GmBH,, Sexau, Germany) with embedded position encoders providing velocity and position feedback. To ensure the drive is capable of sub-millimeter positioning accuracy, the drive was engineered with two gearing stages. Firstly internal to the motors are 19:1 planetary gearboxes, with an average backlash under no load of 1.2°. A second stage involves a worm wheel drive to an anti-backlash gear with a 60:1 reduction. With this final gear connected directly to the wheel (∅40mm), the final play in the system equates to a theoretical positioning ability of 0.006mm. Secondly, with a wheel axle separation of 24mm, a theoretical forward angle alignment of 0.014° is achieved.

Unfortunately greater inaccuracies in the system lie in two fundamental areas, the ability of the wheels to maintain traction on the skull, and in the position feedback of the robot and patient tracking. The second inaccuracy relies on the Polaris Optical Tracking System (Northern Digital Inc., Ontario, Canada), which has known inaccuracies of 0.4mm. The first inaccuracy the traction; however, is more complex and the larger unknown to the system.

The ability of the robot to maintain traction on the skull is unfortunately open to a great deal of unknowns, the skulls curvature or anomalies or contaminates (eg. water, blood or other fluids). Prior to modeling these factors, it was decided to make an overall conceptual test of the friction available from two possible wheel models, shown in Fig. 2: design 1 with an ability to penetrate through any possible surface contaminates, and design 2 for possibly softer, thinner bone (eg children) with a safety flange to prevent complete penetration should too much pressure be placed on the wheels. The result of the testing is discussed in the next section.

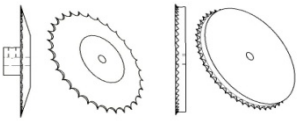


Fig. 2. Two separate wheel designs were prepared for testing on a Swine skull. Design 1 (left) had 2mm spikes, Design 2 (right) had 1mm spikes, with a safety flange to prevent excessive penetration.

2.2 Steering Control System

The required control system here appears similar to that of the collaborative robots (COBOTs) in the work by Peshkin et al. [13]. However, Peshkin defines COBOTs as intrinsically passive devices using “steerable” non-holonomic joints. Instead the Craniostar was designed with driven wheels, providing steering and additionally a greater level of support with the surgeon’s choice of speed, assisting in the forward movement of the device within a set tilt limitation. This tilt limitation also ensures that the cutting tool is used only when very close to vertical; these capabilities would not have been possible through the application of the COBOT control methods. Instead dynamic path following control methods were analysed from a number of papers e.g. [[14],[15],[16]]. The choice was made to test the control algorithm proposed by Seo K. et al [15] for a number of reasons.

- Firstly, the approach does not assume perfect system performance and is designed to cope with uncertainty in dynamics arising from path following.
- Secondly, the approach is not restricted to constant path curvature.
- Thirdly, the algorithm provides the mechanism for definition of a safety boundary. In the initial definition, this safety boundary is an area where the mobile robot may move free from collisions; however, when employing the algorithm in a medical application, the safety boundary is redefined as the area within which it is 'safe' to operate.

A lot of research into unicycle kinematics of mobile robots, is based on the movement of the wheels over a flat surface defined in terms of x and y coordinates, for example see [[14]-[16]]. Fig. 3 *(left)* shows some of the standard parameters for a unicycle robot. Unfortunately this simplistic 2D view of the surface is not perfectly compatible with highly curved and irregular surfaces (eg a skull). An initial experimentation with translation to spherical coordinates was made on a standard model phantom skull. It was determined that the simplistic solution was feasible; however, on complexities that arise out of abnormal skull shapes, such as those found in surgeries requiring birth defect or accident trauma cranial reconstruction, this did not work.

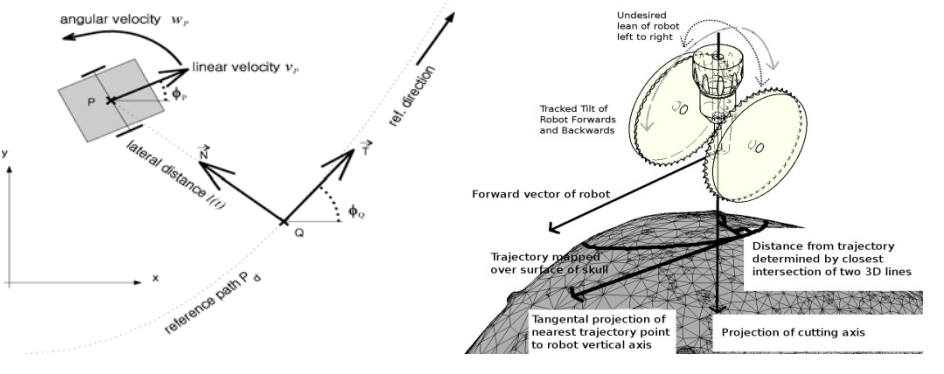


Fig. 3. Original 2D parameters *(left)*(Duplicated with permission from Seo K.) and the 3D derivation *(right)* of the parameters for the Craniostar involving the projection of virtual vectors through the cutting axis of the robot and the closest trajectory segment to this vector.

Instead, we used an approach that entirely ignored the surface. All control parameters were extracted from abstract virtual 3D vectors based on the position of the robot, and the trajectory. Equations to determine the control parameters, discounted the actual intersection with the skull, and for example, looked more directly at the 3D closest segment between two lines such as the closest trajectory segment, and a virtual projection through the drilling axis. The trajectory is known to lie on the skull, and it is known the robot operates with both wheels on the skull, therefore, it can be assumed that the closest line between these two aforementioned vectors will closely approximate the lateral distance on the skull. Fig. 3 displays some of these abstractions. These parameters were then employed in accordance with [15], the key equations being duplicated here: the steering controller in Equation 1, and the sliding mode between safety regions defined in Equation 2. The parameters are defined as per Fig 3, with tuneable parameters k, b_1, and ϵ defining the path approach.

$$w^*(s(t),l(t),\tilde{\phi}(t),v_p(t))=\frac{\kappa(s(t))v_p(t)\cos(\tilde{\phi}(t))}{1-\kappa(s(t))l(t)}-\frac{kv_p(t)\sin(\tilde{\phi}(t))}{1+(kl(t))^2}-(\omega_p+b_1)sat(\frac{z(t)}{\epsilon}). \quad (1)$$

$$z(t)=\tilde{\phi}(t)+\arctan(kl(t))-2\pi m, m\in\mathbb{R}. \quad (2)$$

3 Results

The testing of the Craniostar robot to meet the system requirements, in terms of accuracy, feasibility and safety has commenced in two separate phases.

- The accuracy of the control system was tested initially on flat surfaces and then on phantom skulls, and
- The feasibility and safety aspects were tested under OR conditions on Swine skulls. See Fig. 1.

The tests on the accuracy of the control system are outlined in table 1. Two known inaccuracies lead to less than 100% of the accuracies under the ±0.5mm region. The first in the initial placement of the robot, while the software's visual GUI can guide the surgeon in its placement, it is still found quite difficult to place the robot truly precisely. The recovery from this initial placement occurs quickly, within approximately 2mm of travel. But it is accepted that the initial hole size for inserting the Craniotomy Tool Piece with Dura Guard is larger than this inaccuracy, and is therefore deemed negligible. This is seen in Fig. 4(a) where an initial placement offset was quickly recovered.

Table 1. Average accuracies achieved by the Craniostar on different trajectories

		Average accuracy achieved	
Surface	Trajectory	±0.5mm	±1mm
Flat Wood	Single straight 5cm segment	97.20%	100.00%
	Single curved 90° segment with 4cm radius	98.15%	100.00%
	Two 5cm segments joined with 45° join	97.60%	100.00%
Plastic Phantom Skull	Single straight 5cm segment	97.00%	100.00%
	Two 5cm segments joined with 45° join	95.00%	100.00%
	Single curved 90° segment with 4cm radius	95.60%	100.00%

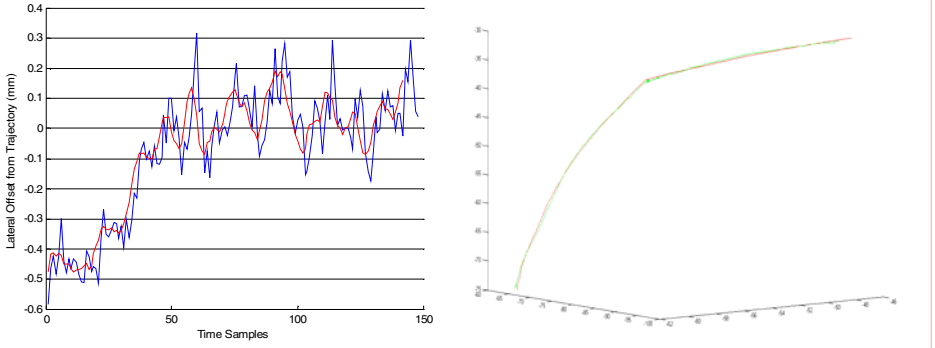


Fig. 4. Tracking examples. (a) The lateral distance (error) of a 5cm straight track showing an initial placement error of the Craniostar, and the following robots correction. Noise in graph is primarily due to Optical Tracking. (b) The 3D Track plot of two joined segments, curving over the surface of the phantom skull (red line, planned trajectory, green line tracked, distances mm)

The second error lies in the nature of the shared control of the system. The hand-held device is by nature prone to noise, specifically in external forces from the surgeons hand. Fig. 4(b) show a trajectory being tracked by a Surgeon over a skull. The intersection between the two segments has a slight disturbance in the middle due to the requirement for sharp turning at the segment intersection. In order to achieve the desired ±0.5mm here, took a degree of education for the user into knowing what the robot is likely to do, this required enhancements to the Graphical User Interface, feeding back to the Surgeon to give advanced warning of such likely intent.

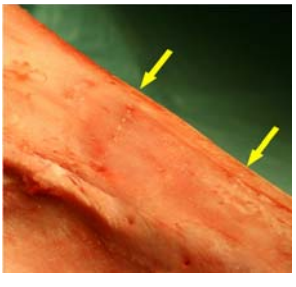


Fig. 5. The superficial marks of the Craniostars motion over the swines skull

The testing on a Swine skull in phase two involved the impact on the skull and frictional testing. The results of this testing demonstrated no ability of the spiked wheels to damage the skull, though superficial tracks did remain on the skull immediately following the robots movements, visible in Fig. 5. The testing also confirmed the robot's ability to move over the skull without any wheel slippage, thus supporting the choice of control theory with non-sliding kinematics.

4 Conclusion

In this paper, a new concept of surgical robotics was proposed and developed using a hand-held mobile robot to aid the surgeon in the conduct of a craniotomy. It shows the possibility for achieving precision pre-surgical planning to OR transfer, without removing the Surgeon from the procedure, and without a large impact on OR footprint and surgical workflows. Future work still includes minaturisation of the gearing section and development of a more precise friction model of the skull for refining the control system prior to moving towards clinical trials.

Acknowledgment. This EU project "CompuSurge" is funded by the "Marie Curie" research network. A collaboration between the University Hospital Heidelberg and the Institute of Process Control and Robotics at Karlsruhe University.

References

1. Hassfeld, S., Muehling, J.: Computer assisted oral and maxillofacial surgery—a review and an assessment of technology. Int. J. Oral Maxillofac. Surg. 30, 2–13 (2001)
2. Weihe, S., Schiller, C., Rasche, C., Hassfeld, S., Wehmöller, M., Knoop, H., Epple, M., Eufinger, H.: CAD-CAM prefabricated individual skull implants: new aspects in robot resection and guided bone regeneration. International Congress Series 1268, 584–590 (2004)
3. Engel, D., Raczkowsky, J., Wörn, H.: RobaCKa- Ein Robotersystem für den Einsatz in der Chirurgie. In: Rechner- und sensorgestützte Chirurgie, Proceedings zum Workshop, vol. 4, pp. 279–286 (2001)
4. Bast, P., Popovic, A., Wu, T., Heger, S., Engelhardt, M., Lauer, W., Radermacher, K., Schmeider, K.: Robot- and computer-assisted craniotomy: resection planning, implant modelling and robot safety. Int. J. Med. Robotics. Comput. Assist. Surg. 2, 168–178 (2006)
5. Nathoo, N., Cavusoglu, M., Vogelbaum, M.A.: In touch with robotics: neurosurgery for the future. Neurosurgery 56, 421–433 (2005)
6. Taylor, R.H., et al.: A model-based optimal planning and execution system with active sensing and passive manipulation for augmentation of human precision in computer-integrated surgery. In: 2nd International Conference on Experimental Robotics, Toulouse, France (1991)
7. Matjaz, J., Harris, S., Baena, F.R., Gomes, P., Davies, B.L.: The Acrobot system for total knee replacement. Industrial Robot: An International Journal 30, 61–66 (2003)
8. Schneider, O., Troccaz, J.: A Six-Degree-of-Freedom Passive Arm with Dynamic Constraints (PADyC) for Cardiac Surgery Application: Preliminary Experiments. Comp. Aid. Surg. 6, 340–351 (2001)
9. Brisson, G., Kanade, T., DiGioia, A., Jaramaz, B.: Precision Freehand Sculpting of Bone. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 105–112. Springer, Heidelberg (2004)
10. Follman, A., Little, J., Schröder, K., Forff, A., Engelhardt, M., Radermacher, K.: Synergistic Control - Semiautomatiche Navigation eines neuartigen Trepanationsinstrumentes. In: CURAC 2008, pp. 111–112 (2008)
11. Kane, G., Eggers, G., Boesecke, R., Raczkowsky, J., Wörn, H., Marmulla, R., Mühling, J.: Intuitively Controlled Handheld Mobile Robot for Precision Craniotomy Surgery. In: 13th International Conference on Human-Computer Interaction (in print, 2009)
12. Schorr, O., Münchenberg, J., Raczkowsky, J., Wörn, H.: KasOp - A Generic System for Pre- and Intraoperative Surgical Assistance and Guidance. In: 15th International Congress and Exhibition of Computer Assisted Radiology and Surgery, CARS (2001)
13. Wannasuphoprasit, W., Gillespie, R.B., Colgate, J.E., Peshkin, M.A.: Cobot Control. In: International Conference on Robotics and Automation (1997)
14. Samson, C.: Path following and time-varying feedback stabilization of a wheeled mobile robot. In: International Conference on Advanced Robotics and Computer Vision, vol. 13, pp. 1–5 (1992)
15. Seo, K., Lee, J.S.: Kinematic path-following control of a mobile robot under bounded angular velocity error. Advanced Robotics 20, 1–23 (2005)
16. Kim, B., Tsiotras, P.: Controllers for Unicycle-Type Wheeled Robots: Theoretical Results and Experimental Validation. IEEE Transactions on Robotics and Automation 18, 294–307 (2002)

Dynamic Active Constraints for Hyper-Redundant Flexible Robots

Ka-Wai Kwok[1], George P. Mylonas[1], Loi Wah Sun[1], Mirna Lerotic[1], James Clark[1], Thanos Athanasiou[2], Ara Darzi[3], and Guang-Zhong Yang[1]

[1] Royal Society/Wolfson Medical Image Computing Laboratory, Institute of Biomedical Engineering
[2] Department of Bio-Surgery and Surgical Technology
[3] Department of Surgical Oncology and Technology,
Imperial College London, London, United Kingdom
{k.kwok07,george.mylonas,l.sun,mirna.lerotic,j.clark, t.athanasiou,a.darzi,g.z.yang}@imperial.ac.uk

Abstract. In robot-assisted procedures, the surgeon's ability can be enhanced by navigation guidance through the use of virtual fixtures or active constraints. This paper presents a real-time modeling scheme for dynamic active constraints with fast and simple mesh adaptation under cardiac deformation and changes in anatomic structure. A smooth tubular pathway is constructed which provides assistance for a flexible hyper-redundant robot to circumnavigate the heart with the aim of undertaking bilateral pulmonary vein isolation as part of a modified maze procedure for the treatment of debilitating arrhythmia and atrial fibrillation. In contrast to existing approaches, the method incorporates detailed geometrical constraints with explicit manipulation margins of the forbidden region for an *entire* articulated surgical instrument, rather than just the end-effector itself. Detailed experimental validation is conducted to demonstrate the speed and accuracy of the instrument navigation with and without the use of the proposed dynamic constraints.

1 Introduction

Robotic-assisted minimally invasive cardiac surgery has been the major aim of many robotic platforms. The tight confines of the thoracic cavity and mediastinum however, challenge even the most skilled surgeons as limited field-of-view and restricted maneuverability command advanced manual dexterity and hand-eye coordination. These difficulties become even more evident during beating heart surgery, raising critical issues with respect to procedural safety and precision. This is of particular concern during procedures such as minimally invasive bilateral pulmonary vein isolation performed as part of the modified maze approach, to the treatment of chronic and paroxysmal atrial fibrillation when the heart is not only in motion but also beats irregularly. The maze procedure describes the transmural epicardial ablation of a complete circle around the pulmonary veins, an area of the heart known to generate aberrant arrythmogenic electrical activity central to the pathophysiology of atrial fibrillation. Currently, the treatment of atrial fibrillation is through an endovascularly approached endocardial ablation or during open heart surgery with poor response rates [1]. The

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 410–417, 2009.

maze procedure is more effective but requires either a median sternotomy or multiple small incisions on both the left and right sides of the chest to ensure that complete encirclement of the vessels has been correctly affirmed [2]. For access via a minimally invasive approach, a catheter is required to pass through two narrow spaces posterior to the heart, tightly bordered by the great vessels. These spaces are roofed by the approximated pericardium. The challenge of ensuring the path for the catheter is accuracy, reflecting upon both the lack of visual field and maneuverability within these spaces. The risk of accidental instrument injury is hampering its uptake despite recognized beneficial patient outcomes.

In robot-assisted procedures, the surgeon's ability can be enhanced by augmented manipulation guidance such as virtual fixtures or active constraints to enhance human sensory feedback and limit aberrant instrument maneuvers [3, 4]. Such concepts have been used to guide the cutting of the tibia and the removal of the blockages within a permitted region in orthopedic knee surgery [5] and endoscopic sinus surgery [6], respectively. They enable the operator to manipulate the tool tip along the desired 3D path without applying excessive force on the contact surface of delicate surroundings. For complex geometries, spatial motion constraints have to be adapted in real-time, making them suitable for surgical interventions. Previous work has so far mainly relied on the preoperative segmented image data or an anatomical model. Moreover, the constraints are assumed to be static throughout the operation. This, however, is not practical especially in cardiac procedures, which involve large respiratory and cardiac induced tissue deformation. The future clinical impact of active constraints or virtual fixtures requires not only flexibility in updating the constraints, but also the ability to react to changes in tissue morphology intra-operatively. Recently, virtual fixtures have addressed dynamic surgical scenes where tissue deformation is present. For example, Ren *et al*. [7] proposed dynamic 3D virtual fixtures for beating heart ablation procedures. The work complements manipulation guidance with intra-operative sensing data. The dynamic virtual fixture model is defined based on the pre-operative data, then registered to intra-operative images. Most of the existing methods, not excluding [6], are concerned with confining the motion of a single-point end-effector rather than the instrument body, even if it is rigid. Current research into robotic-assisted cardiac surgery has been focused on developing more flexible hyper-redundant devices to overcome the problem of safe access and navigation within the tight confines of the chest cavity and pericardial space [8, 9]. Deformation of the heart can have a profound effect on spatial constraints. To overcome these difficulties, dynamic active constraints could permit rapid and safe transthoracic and intrapericardial navigation without injuring the surrounding tissue or organs.

The purpose of this paper is to introduce a real-time modeling scheme for dynamic active constraints with adaptation to cardiac deformation and anatomic structure changes. A smooth and dynamic cylindrical pathway with detailed geometrical constraints defines the explicit manipulation margin of the forbidden region for an entire articulated surgical instrument (*e.g.* a snake robot) rather than just the end-effector itself. Detailed experimental validation is conducted to demonstrate the speed and accuracy of the instrument navigation with and without the use of the proposed dynamic constraints.

2 Method

2.1 Construction of Dynamic Active Constraints

Dynamic active constraints are pre-determined constraining pathways that adjust to tissue deformation in real time. This method helps the surgeon by constraining the motion within a safety margin for a prescribed path rather than discrete points. First, an obstacle-free centerline is generated as a collection of reference control points that describe the convoluted path. By using a 3D coordinate input device, the operator can place a series of control points which determine a space curve. Because of manual point placement, a refined adjustment and increment of points is conducted automatically by checking collision with the anatomical model.

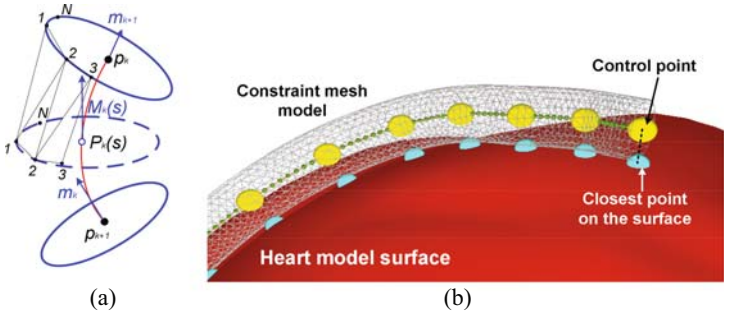


Fig. 1. (a) Triangular mesh formed between two adjacent rings of the model for active constraints; (b) the constraint represented as a tube-pathway mesh boundary along the parametric centerline

Each control point can be expanded spherically in 3D space, *i.e.,* a margin is defined by the control radius of a sphere; considering the dimension of the adopted articulated instrument. If the distance to the epicardial surface is larger or smaller than the margin, the locations of the control points are readjusted towards the centre of the sphere. When the instrument manipulation is to follow a path within a prescribed boundary, a parametric centerline is defined as $C(s)$: $s \in [0,1] \rightarrow \Re^3$. Hermite curve (*cspline*) is used in this study to represent the centerline determined by N_c control points. The advantage of using *cspline* over other types of approximation is that it provides intuitive geometric manipulation for adjusting a set of parameters by which the generated curve can be adapted to the dynamic environment. For interpolation with the control points, the polynomial curve segment C_k at $s \in (0,1)$ is defined as

$$C_k(s) = h_{0,0}(s)p_k + h_{1,0}(s)m_k + h_{0,1}(s)p_{k+1} + h_{1,1}(s)m_{k+1} \quad (1)$$

where $k = 1,\ldots,N_c\text{-}1$, $h_{0,0}$, $h_{1,0}$, $h_{0,1}$ and $h_{1,1}$ are 3rd-order polynomial functions with respect to s. p_k and p_{k+1} are the start and end points of the curve and m_k and m_{k+1} are the corresponding tangential values at these points, where $m_1 = p_2 - p_1$, $m_{Nc} = p_{Nc} - p_{Nc\text{-}1}$ and $m_k = (p_{k+1} - p_{k\text{-}1})/2$, for $k \neq 1$ and $k \neq N_c$. Tangent $M_k(s)$ at intermediate point $C_k(s)$ is

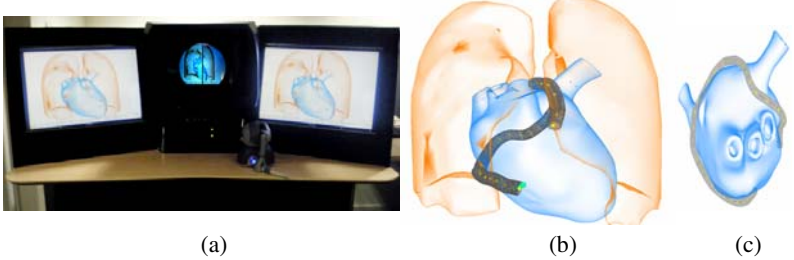
interpolated as $M_k(s)=(1-s)m_k+s(m_{k-1})$. Fig. 1(a) illustrates the rings centered at $C_k(s)$ and their orientations conformance to $M_k(s)$. Each ring is represented by a polygon of radius r. N number of polygon vertices are aligned in an anti-clockwise direction. Before being transformed along the curve, the polygon vertices are positioned on the X-Y plane and centered at the origin. The coordinates of the vertices, X_i, are determined by Eq. (2). To align the Z-axis of the polygonal coordinate frame parallel to $M_k(s)$, the equivalent rotation matrix $R(\kappa,\theta)$ has to be computed, which denotes the rotation about a unit vector κ by an angle of θ calculated in Eq. (3). Therefore, the Cartesian coordinates of polygon vertex X'_{ki} relative to the global coordination frame can then be obtained in Eq. (4), such that the vertices are centered at point $C_k(s)$ and along the curve segment k.

$$X_i = r\left[\sin\left(2\pi i / N\right), \cos\left(2\pi i / N\right), 0\right]^T \tag{2}$$

$$\kappa_k(s) = norm([0,0,1]^T \times M_k(s)) \quad \text{and} \quad \theta_k(s) = \cos^{-1}\left(M_k(s)\cdot[0,0,1]^T / \left|M_k(s)\right|\right) \tag{3}$$

$$X'_{ki}(s) = R\left(\kappa_k(s), \theta_k(s)\right) X_i + C_k(s) \tag{4}$$

The algorithm for mesh generation interconnects the vertices between two adjacent rings, thus forming a triangular mesh structure. To provide a smooth boundary, collision checking between two adjacent rings is performed. If the two adjacent rings are found to be colliding with each other, the interconnection will be formed by choosing another ring with larger separation. As shown in Fig. 1(b), a smooth cylindrical pathway describing the active constraint is constructed. Since only point-to-object collision checking is involved, the adjustment is fast and can be conducted in real time. The objective of active constraints is to confine the instrument manipulation with the margins of the forbidden region being extended cylindrically along the centerline. The cylindrical radius is a variable that can be adjusted to allow for movement of the robot within the tunnel. It is determined by the dimension of the flexible robot and the required volumetric tolerance. For this study, it was set to 2mm for all experiments. Furthermore, the upper constraints can extend beyond the pericardium to accommodate stretching.



(a) (b) (c)

Fig. 2. The 3D auto-stereoscopic environment used for surgical navigation: (a) front view; The dynamic active constraints generated on heart model: (b) shown with the lung model and (c) viewed from top

2.2 Experimental Setup

A high resolution 3D auto-stereoscopic display system (IRIS-3D, UK) was used to provide a navigation environment as shown in Fig. 2(a). The operator can view stereo images while navigating through the anatomical pathways. A synthetic simulation environment with a deforming heart and lung model was created. The 3D model was reconstructed from the CT data of a cardiac phantom (Chamberlain Group, MA, USA). The method was validated by recruiting nine subjects to assess the performance of the dynamic active constraints. Subjects were asked to operate a phantom tool device (Omni Phantom, SensAble Tech. Inc., USA) and trace a desired path on the beating heart model. To simulate ablation pathways similar to a maze procedure, dynamic active constraints were constructed and wrapped around the heart as shown in Fig. 2. To simulate realistic cardiac motion, interpolation was used between prescribed key frames including static heart reference H, deforming long L and short S cardiac axes. These were superimposed as described by Eq. (5).

$$V(t) = V_H + \left(V_S - V_H\right)\frac{\sin(2\pi f_S t)+1}{2} + \left(V_L - V_H\right)\frac{\sin(2\pi f_L t)+1}{2} \qquad (5)$$

where V_H, V_L and V_S denote the vertices location at the key frames. In this experiment, the two modes of deformation were set to repeat at a rate of f_S Hz and f_L Hz, with f_S=2f_L (*e.g.* f_S =1, 60 bpm (*beats per minute*)) in order to provide realistic periodic motion. In all experiments, 60 bpm was chosen as the reference heart rate as the average heart rate of a healthy adult at rest is around 60 to 80 bpm. Subjects were required to operate the surgical tool tracing a desired ablation path on the heart model with (60 bpm) and without cardiac motion. Each subject performed the task twice by using free-hand manipulation and constrained by forces generated by the phantom device. The order of the permutation was randomized to prevent bias due to learning effects. Six performance indices were recorded. They include task completion time, the maximum and average depth of collision onto the heart model, the number of collisions recorded, the total distance travelled during collision (collision path length) and the average of the total distance when the tool deviated from the desired trajectory (path deviation). Data analysis was performed by using a between-group comparison statistical test (One-Way ANOVA with Bonferroni test, SPSS Inc., Chicago, IL). Four groups were compared: 1) Static heart without constraints; 2) Static heart with constraints; 3) Heart beating at 60 bpm without constraints; 4) Heart beating at 60 bpm with constraints. Each of them is mutually independent among the six performance indices. The main purpose of the performance metrics is to demonstrate the confidence gain by the use of dynamic active constraints. These metrics are related to the deviation of the instrument from the ideal path. Instrument-tissue forces were not rendered in the case of free-hand operation. In order to assess how the heart rate affects the accuracy of manipulation, the experiment as described above was also conducted with the heart beating rate varying from 60 to 120 bpm.

3 Results

Using multiple comparison tests, statistically significant differences (α<0.05) occurred between the four groups among all six performance indices. More specifically,

obtained α-values ranged within [0, 0.008]. The mean and standard deviation of the six performance indices for all subjects, with and without the use of constraints, are summarized in Table 1. It is evident that with the introduction of dynamic active constraints, the overall performance has been improved significantly. For both static and dynamic experiments, improved scores are obtained for all indices when active constraints are activated. No collision was found when subjects traced a desired path on the static heart model with constraints. Overall, navigation with constraints has shown reduced task completion time and path deviation. Furthermore, this performance gain seems to be independent of the static or dynamic environment, showing less than 1% variance. Fig. 3(a) illustrates the path deviation for one of the subjects studied. It can be seen that path deviation with constraints is less than that without constraints, maintained at a steady level of 2mm. This was attributed to the fact that the subject followed the inner surface of the pathway during operation. With the use of constraints, the reduced path variation reflects the confidence of the operator during 3D maneuver. Fig. 3(b) shows that the path-to-path variance of path deviation for the entire experiment was less than 1mm. Unsurprisingly, the trajectory of the tool tip while performing path following with constraints is close to the desired path as shown in Fig. 3(c). These experiments were repeated with a range of heart rates (60 to 120 bpm) and the corresponding results are shown in Table 2, where the values shown are normalized performance gains for each index with the use of constraints.

Standardization of data was achieved at different frequencies from a single subject. In Table 2, the percentage change was calculated by using the performance indices of free-hand operation (*i.e.*, without the use of active constraints) as the reference. It is evident from these results that the overall performance gain is maintained as the heart rate increases. During high-frequency motion, the overall hand-eye coordination is generally deteriorated, highlighting the need of virtual motion stabilization in these cases. To better visualize the change in the margins associated with these gains, Fig. 4

Table 1. Summary of changes in performance indices with and without the use of dynamic active constraints (averaged across the nine subjects studied)

	Without constraints				With constraints			
	Static heart		Beating heart (60 bpm)		Static heart		Beating heart (60 bpm)	
	mean	SD	mean	SD	mean	SD	mean	SD
Completion time (sec)	35.9	(15.3)	37.6	(16.6)	12.9	(10.1)	11.0	(5.1)
Max collision depth, mm	14.6	(8.5)	15.0	(11.8)	0.0	(0.0)	0.7	(0.9)
Mean collision depth, mm	7.5	(4.8)	6.8	(5.6)	0.0	(0.0)	0.7	(0.8)
No. of collisions	5.0	(2.4)	7.0	(2.6)	0.0	(0.0)	0.7	(0.9)
Collision path length, mm	123.8	(59.7)	112.0	(90.9)	0.0	(0.0)	2.9	(3.9)
Path deviation, mm	8.4	(4.6)	12.8	(9.4)	2.0	(0.0)	1.7	(0.3)

Table 2. Percentage change in performance indices for varying heart rate (60-120 bpm)

Heart rate (bpm)	60	70	80	90	100	110	120
Completion time (%)	60.1	60.3	59.0	45.3	66.7	76.8	50.1
Path deviation (%)	23.9	84.6	42.9	34.0	58.3	63.8	31.0

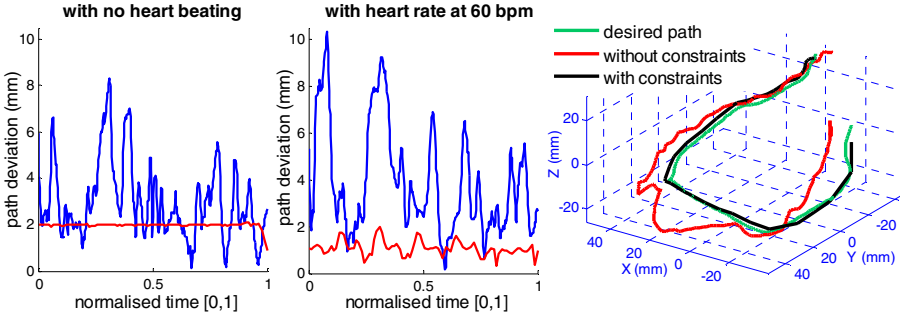


Fig. 3. Example path deviation plots for one of the subjects studied, showing the absolute path deviation and variance throughout the procedure with and without constraints when the heart is static (a) and beating (b). (c) Trajectory of the virtual tool tip in following the desired path.

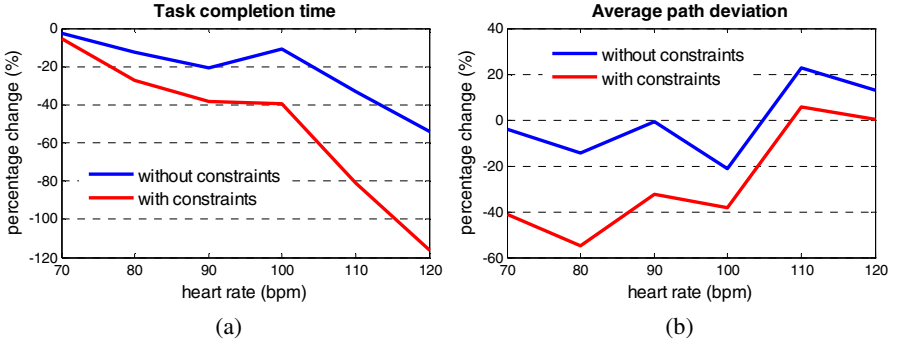


Fig. 4. Percentage change in task completion time (a) and average path deviation (b) with and without the use of constraints normalized against the respective indices at 60 bpm

illustrates the trend of the task completion time and average path deviation with increasing heart rate. In this figure, the percentage change is calculated by using the respective performance indices at 60 bpm as the reference. In Fig. 4(a), a decrease (more negative) in task completion time implies that shorter time is required to complete the task for both cases. Similarly, increased path-following errors are indicated by the percentage increase of path deviation as shown in Fig. 4(b). This can be attributed to the fact that motor tracking is worse while the desired path is moving faster at higher heart rates. As a result, the subject was not able to perform detailed tracking of a particular region.

4 Discussion and Conclusions

In this paper, a real-time modeling scheme for dynamic active constraints is proposed. Validation through the use of a simulated beating heart model demonstrates its ability in adapting to changes of tissue morphology. Experimental validation has shown

statistically significant improvement in terms of speed, accuracy and minimized tissue damage when dynamic active constraints are introduced. The clinical relevance of the study is that in many cardiac intervention procedures, particularly the maze procedure, it is not appropriate to use the pericardium and the heart itself as a physical constraint to limit the motion of the robotic probe normal to the epicardial surface. Forces normal to the epicardial surface can penetrate the myocardium injuring coronary vessels or perforating the pericardium and damaging sensitive nerves such as the phrenic nerve. In this case, flexible, non-articulated devices may not provide accurate 3D navigation. This has motivated the development of hyper-redundant flexible robots with 3D dynamic active constraints. The proposed motion constraints concept can readily incorporate manipulation margin of long articulated surgical instruments rather than just the end-effectors. Such a scheme is also suitable for integrating other navigation schemes, *e.g.*, gaze contingent motor channeling [10], for further enhancing the surgeon's hand-eye coordination.

Acknowledgments. The work described in this paper was supported by grants from the UK EPSRC (DT/E011101/1, EP/D057213/1), TSB and the Welcome Trust.

References

1. Hornero, F., Rodríguez, I., Bueno, M., Buendía, J., Dalmau, M.J., Canovas, S., Gil, O., Garcia, R., Montero, J.A.: Surgical ablation of permanent atrial fibrillation by means of maze radiofrequency: mid-term results. J. Card. Surg. 19, 383–388 (2004)
2. Raman, J.S., Seevanayagam, S., Storer, M., Power, J.M.: Combined endocardial and epicardial radiofrequency ablation of right and left atria in the treatment of atrial fibrillation. Ann. Thorac. Surg. 72, 1096–1099 (2001)
3. Rosenberg, L.B.: Virtual fixtures: Perceptual tools for telerobotic manipulation. In: Proc. of the IEEE Annual International Symposium on Virtual Reality, pp. 76–82 (1993)
4. Okamura, A.M.: Methods for haptic feedback in teleoperated robot-assisted surgery. Industrial Robot: An International Journal 31, 499–508 (2004)
5. Davies, B., Jakopec, M., Harris, S.J., Baena, F.R.Y., Barrett, A., Evangelidis, A., Gomes, P., Henckel, J., Cobb, J.: Active-constraint robotics for surgery. Proc. of the IEEE 94, 1696–1704 (2006)
6. Li, M., Ishii, M., Taylor, R.H.: Spatial motion constraints using virtual fixtures generated by anatomy. IEEE Trans. Robotics 23, 4–19 (2007)
7. Ren, J., Patel, R.V., McIsaac, K.A., Guiraudon, G., Peters, T.M.: Dynamic 3-D virtual fixtures for minimally invasive beating heart procedures. IEEE Trans. on Medical Imaging 27, 1061–1070 (2008)
8. Zenati, M.A., Bonanomi, G., Chin, A.K., Schwartzman, D.: Left heart pacing lead implantation using subxiphoid videopericardioscopy. J. Cardiovasc Electrophysiol. 14, 949–953 (2003)
9. Zenati, M.A., Shalaby, A., Eisenman, G., Nosbisch, J., McGarvey, J., Ota, T.: Epicardial left ventricular mapping using subxiphoid video pericardioscopy. Ann. Thorac. Surg. 84, 2106–2107 (2007)
10. Mylonas, G.P., Kwok, K.W., Darzi, A., Yang, G.-Z.: Gaze-contingent motor channelling and haptic constraints for minimally invasive robotic surgery. In: Proc. of Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, vol. 11, pp. 676–683 (2008)

Nonmagnetic Rigid and Flexible Outer Sheath with Pneumatic Interlocking Mechanism for Minimally Invasive Surgical Approach

Hiromasa Yamashita[1], Siyang Zuo[1], Ken Masamune[1], Hongen Liao[2], and Takeyoshi Dohi[1]

[1] Graduate School of Information Science and Technology, The University of Tokyo, Japan
[2] Graduate School of Engineering, The University of Tokyo, Japan

Abstract. We developed a nonmagnetic rigid and flexible outer sheath with pneumatic interlocking mechanism using flexible toothed links and a wire-driven bending distal end. The outer sheath can be switched between rigid and flexible modes easily depending on surgical scenes, and the angle of its distal end can be controlled by three nylon wires. All components of flexible parts are made of MRI-compatible nonmagnetic plastics. We manufactured the device with 300-mm long, 16-mm outer diameter, 7-mm inner diameter and 90-mm bending distal end. Holding power of the device in rigid mode was maximum 3.6 N, which was sufficient for surgical tasks in body cavity. In vivo experiment using a swine, our device performed smooth insertion of a flexible endoscope and a biopsy forceps into reverse side of the liver, intestines and spleen with a curved path. In conclusion, our device shows availability of secure approach of surgical instruments into deep cavity.

1 Introduction

Minimally invasive surgery (MIS) is in favor of closed or local surgery with fewer traumas unlike conventional invasive open surgery. In MIS, endoscopic devices and instruments are usually used to enable remote control surgical procedures with indirect observation of the surgical field through patients' body surface such as abdominal wall and chest wall into their body cavity, or through natural orifice such as mouth, urethra and anus. MIS allows patients to have shorter hospitalization time, outpatient treatment and earlier rehabilitation into society.

Various endoscopic devices and instruments have been researched and developed to improve surgical dexterity, and to reduce specific risks and difficulties in endoscopic surgery. For example a dexterous robotic manipulator added multiple degrees of freedom (DOFs) to operation of the instruments and thus it improved surgical performance [1-3]. So-called master-slave robotic manipulators, such as previously-commercialized da Vinci® Surgical System, enable surgeons to work with precise operation [4]. However, some problems remain unsolved in these surgical robotic systems. First, familiar laparoscopic surgery, which is one of the endoscopic surgeries, requires a sufficient space below the abdominal wall for surgical procedures. A pneumoperitoneum is commonly used to secure the space, but some

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 418–425, 2009.

complications have been reported from this method such as compression of organs by filled gas and solve out of CO_2 gas into vessels. Second, laparoscopic surgery is useful when the affected area can be approached from the anterior of the body, however, it is difficult to approach the target in a deep and narrow area or reverse side of organs. Some research groups have developed flexible manipulators with a large curved radius. Ikuta et al. developed a micromanipulator to approach inaccessible regions [5]. Other flexible manipulators have been developed using shape memory alloys [6] or a wire-driven mechanism [7]. These manipulators can be inserted in the narrow space and approach affected area, however stabilized path and local surgical space can not be secured for other instruments such as a laparoscope, a forceps etc.

To solve this issue, some systems with lockable sheaths or hybrid locking and relaxing function have been developed. Robert et al. applied this idea to medical applications [8]. Snake shaped robot, called HARP, was also developed [9]. Yagi et al. developed a guiding device including an inner channel, which uses flexible manipulators to approach deep regions [10]. These manipulators have some limitations because their mechanisms of distal bending part are very complicated and their diameters must be larger than the size required for laparoscopic surgery. Moreover, it is difficult to make their components nonmagnetic material to realize MRI-compatibility. MRI-guided surgery is effective for such as minimally invasive therapy of tumors throughout the body [11].

On inserting an instrument into the area of interest without any damages to around tissues, surgical instruments require special function to switch a flexible mode for free insertion and a rigid mode for stabilizing the outer sheath. In this study, we developed a rigid and flexible outer sheath with pneumatic locking mechanism using flexible toothed links and with wire-driven active-bending distal end for laparoscopic surgery.

This paper reports on (1) the mechanism switching between a flexible and a rigid modes, (2) a prototype of the outer sheath device with an active-bending distal end and (3) mechanical performance evaluation of the device. Moreover, we evaluated it (4) in vivo experiment using a swine.

2 Materials and Methods

2.1 Outer Sheath Design

The outer sheath provides a secure approach path for the other surgical instruments inside the human body. The sheath can be flexible and twisted into any shape passively and can still retain its shape against external forces or insertion frictional force from inserted instruments. Before inserting flexible surgical instruments, the surgeon inserts the outer sheath with flexible mode manually through the narrow gap between the safety areas under endoscopic or MRI-guidance. When insertion direction of the outer sheath should be controlled, its wire-driven distal end is bent selectively. After the sheath approaches the target, the surgeon locks the shape and then inserts flexible instruments easily through the internal path made by the sheath.

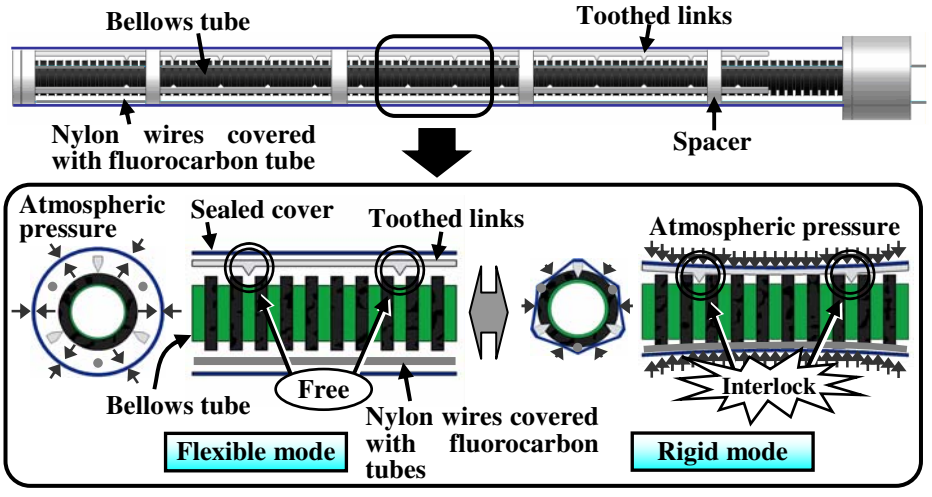


Fig. 1. Switching mechanism of the outer sheath between a flexible and a rigid modes. When there is no vacuum, the toothed links and the bellows tube can take any shape in a sealed clearance space. On the other hand when the pneumatic pump creates a vacuum, the links are pressed and meshed with the ditch of the tube, locking the shape of the sheath.

The outer sheath can be switched between a flexible and a rigid modes easily (Fig. 1). This mechanism consists of three flexible toothed links, a centered bellows tube and a surrounding sealed cover. In the flexible mode, inside air pressure between the bellows tube and the sealed cover is equal to atmospheric pressure. Because there becomes a sealed clearance space between the bellows tube and the sealed cover, teeth of the toothed links disengage from chased external surface of the bellows tube. In the rigid mode, sealed clearance space between the bellows tube and the sealed cover is evacuated by discharging the internal air, and the atmospheric pressure presses the toothed links into the external surface of bellows tube. These interlocked links and bellows tube lock the shape of the outer sheath. And three nylon wires are passing through the clearance space along the bellows tube to bend the distal end of the outer sheath. These wires are covered with unshrinkable fluorocarbon tubes.

2.2 System Configuration

System configuration of the prototype of the outer sheath is shown in Fig. 2. The outer sheath has an outer diameter of 16 mm, inner diameter of 7 mm, length of 300 mm and can achieve a minimum curvature radius of 45 mm. The outer sheath, as described above, consists of three flexible toothed links, a bellows tube, three nylon wires covered with fluorocarbon tubes and a polyethylene sealed cover. The toothed links and bellows tube are flexible, and their shapes are free and changeable. The three toothed links are placed 120 degrees apart, making it possible to lock the outer sheath in any direction in three-dimensional space. The ticks of the bellows tube are 2 mm apart, and the locking teeth on the links are 17 mm apart. All parts of the outer sheath are made of plastic, ensuring MRI-compatibility. The bellows tube is connected to a vacuum pump (DTC-41, ULVAC KIKO INC., Japan) and switching controller (V030E1, Koganei, Japan) to change condition of the outer sheath between

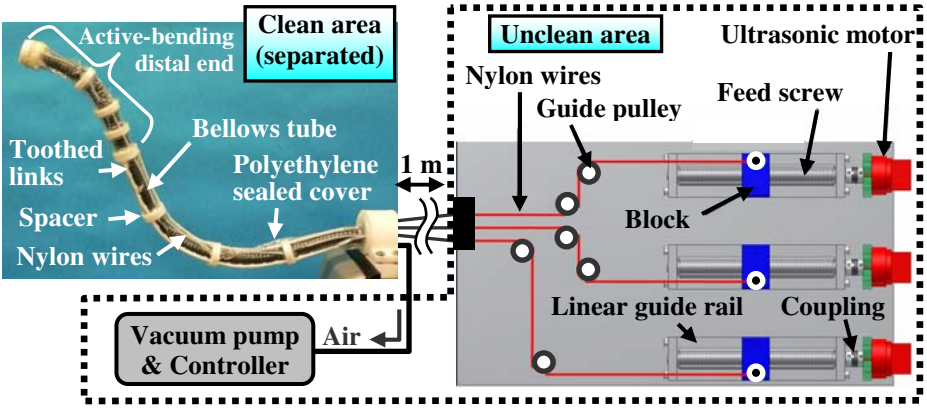


Fig. 2. System configuration of the outer sheath device. Rigid and flexible modes can be switched by the vacuum pump and controller. Bending angle of the distal end is controlled by three sets of ultrasonic motors. The outer sheath is 1-m separated from other parts for cleanness.

a flexible and a rigid modes. In the rigid mode, flexible toothed links engage the bellows tube and they are locked by the vacuum. The shape of the outer sheath can be locked even with a twisted shape. When the vacuum is released, the sheath switches back to the flexible mode. Using this mechanism, it is possible to maintain a rigid shape for the outer sheath whenever a vacuum is continuous. In addition the device has a 90-mm wire-driven distal end. The three nylon wires, which are also placed 120 degrees apart, drive bending angle of this part to arbitrary direction by ultrasound motors (USR30-E3N, Shinsei Corporation, Japan) with couplings and feed screws. Surgeons can control bending angle by rotational three knobs. Materials of the linear-guides are also plastic to realize MRI-compatibility. The outer sheath is specifically separated from the actuators' part and the vacuum pump to be cleanable and sterilizable for clinical use.

2.3 Wire-Driven Bending Mechanism

Each nylon wire is passed through spacers aligned around the bellows tube. When the wire is pulled, length of the wire at one unit between adjacent two spacers is l_i changes to l_i' depending on the bending angle θ_i (**Fig. 3(A)**). Relationship between l_i and l_i' are described by an equation (1) with geometric relations in Fig. 3(A).

$$\frac{l_i}{2} : b\,\frac{\theta_i}{2} = b\sin\frac{\theta_i}{2} + \frac{l_i'}{2} : b\sin\frac{\theta_i}{2} \tag{1}$$

Displacement of the wire Δl_i is, when θ_i is not zero, described by an equation (2) and total displacement with n units $Wire_{pull}$ is described by an equation (3). On the other hand when θ_i is zero, Δl_i and $Wire_{pull}$ are also zero.

$$\Delta l_i = l_i - l_i' = l_i - \left(\frac{2l_i}{\theta_i} - 2b\right)\sin\frac{\theta_i}{2} \quad (2), \qquad Wire_{pull} = \sum_{i=1}^{n} \Delta l_i \tag{3}$$

Relationship between total bending angle and total wire displacement is shown in Fig. 3(B) with 8 units including 4-units distal end of the prototype shown in Fig. 2. Each bending range of one unit is about 22.5 degrees and in case of bending with only the distal end part, maximum bending angle is about 90 degrees. On the other hand bending of the whole device with eight units achieves about 180 degrees.

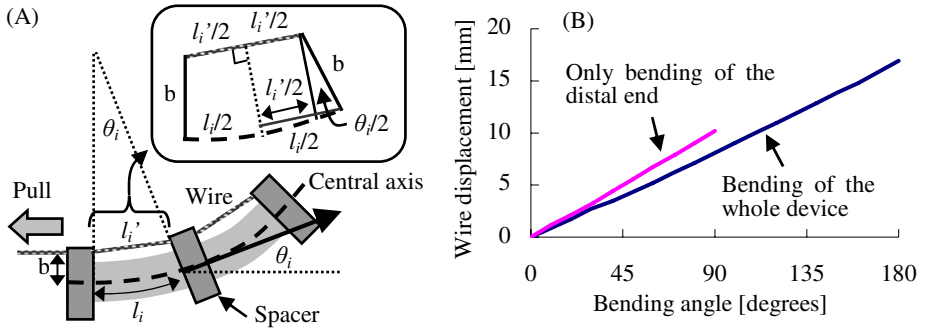


Fig. 3. (A) Wire-driven bending model at one unit between adjacent two spacers. Pulling wire performs each unit's bending motion. (B) Relationship between bending angle and wire displacement of only with the distal end and the whole device.

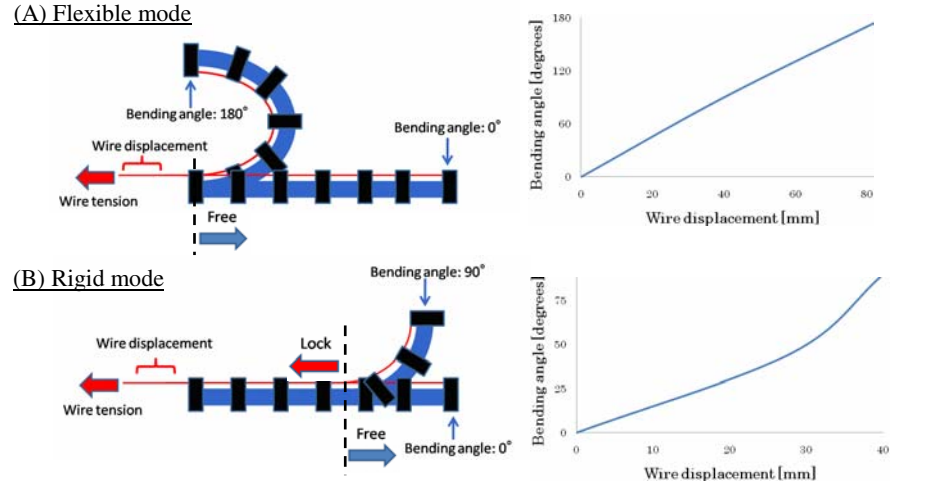


Fig. 4. Active-bending characteristics of the outer sheath by pulling one wire. (A) Result in flexible mode. (B) Result in rigid mode.

3 Results

3.1 Mechanical Performance Evaluations

First, we examined bending characteristics by pulling wire in the flexible and the rigid modes. In the flexible mode, only one wire was used to bend whole outer sheath including the distal end. In the rigid mode, the active-bending distal end was only bent. In the flexible mode, bending angle of 180 degrees was achieved with 85-mm wire displacement and 11.7-N wire tension (Fig. 4(A)). In the rigid mode, bending angle of 90 degrees was achieved with 40-mm wire displacement and 16.6-N wire tension (Fig. 4(B)). Second, we examined limited holding powers of outer sheath as keeping its shape using only the toothed links interlocking mechanism in rigid modes. Limited

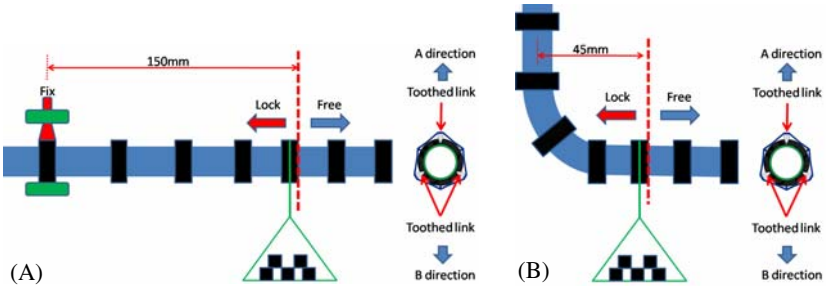


Fig. 5. Measurement of holding power in rigid mode. (A) Load on direction B from direction A in straight condition (0°). (B) Load on direction B from direction A in curved condition (90°).

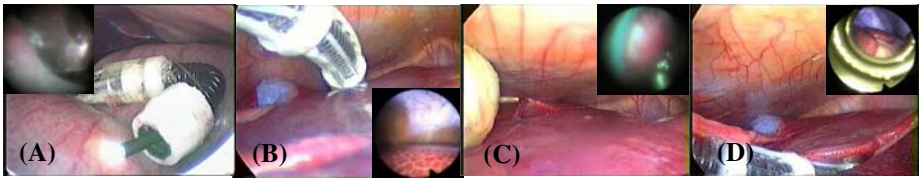


Fig. 6. Laparoscopic broad views with fiberscope local views using our device inserted into abdominal cavity of swine. (A) Colon observation with a widely curved distal end inserting a fiberscope, which shows reverse side of the colon. (B) The outer sheath goes into the gap between liver and abdominal membrane. Fiberscope shows space between liver and abdominal membrane. (C) Liver lobes clamping image by a biopsy forceps inserted the outer sheath. Fiberscope shows clamped liver lobes. (D) The outer sheath goes into the gap between spleen, liver and stomach. Fiberscope shows between their organs.

holding powers were values where the loaded outer sheath's shape started to deform, that is, the toothed links started to be unlocked, and we could observe it by the eye. The outer sheath was fixed at the second spacer from based side and measured in two cases, one was in straight condition (0°) (Fig. 5(A)), and the other was in curved condition (90°) (Fig. 5(B)). Loaded direction was downward vertically in both cases. Maximum load without deformation of the shape were 1.6 N in case (A) and 3.6 N in case (B).

3.2 *In vivo* Experiment

In *in vivo* experiment, we tested usability of the device in the abdominal cavity of a swine (male, 45.5 kg). Inserted instruments into the outer sheath were a 2.8-mm fiberscope (SUMITA Optical glass, Inc., Japan), a 6-mm high resolution fiberscope (VH-F61, KEYENCE CORPORATION, Japan) and a 1.75-mm biopsy forceps (BF1812SF, FUJINON TOSHIBA ES SYSTEMS CO. LTD, Japan). A 2.8-mm fiberscope and a 1.75-mm biopsy forceps could be inserted at the same time. We used a normal laparoscope to observe and confirm position and shape of the device.

First, we tried observation of the colon from the back side with a 2.8-mm fiberscope (Fig. 6(A)). Second, we inserted the device into the gap between liver and abdominal membrane to observe a back side of the liver with a 6-mm fiberscope (Fig. 6(B)). Third, we clamped liver lobes surface by a biopsy forceps with a 2.8-mm

fiberscope view (Fig. 6(C)). Fourth, we inserted the device into the gap between spleen, liver and stomach. These tasks were not supported by any forceps (Fig. 6(D)).

4 Discussion

In mechanical performance evaluations, we confirmed that our device performed large bending angle in both flexible and rigid modes, which enabled switching between 0 ~ ±180-degrees bending of the whole device in the flexible mode and 0 ~ ±90-degrees bending of the distal end of the device in the rigid mode depending on various surgical cases. In this study only a 90-mm distal end could be actively bent in the rigid mode, if the length of this part is variable, our device may enable more dexterous and secure approach to apply to complicated deep cavity. And if the number of active DOF increases, the wire-driven mechanism would become more complex, however its appearance is not so changed because the wires are thin to the outer sheath. More in vivo experiments should be needed to evaluate the suitable constitution of active-bending parts and passive flexible parts for various surgical applications. About holding powers of outer sheath to keep its shape in the rigid mode, it is said that 5-N is required for forceps power in general laparoscopic surgery to raise and hold heavy organs [12]. However our device purposes only insertion of flexible instruments into gap between organs, therefore even 1.58 N in straight shape and 3.58 N in 90-degrees curved shape may be sufficient against external pressure from organs and internal pressure from inserted instruments through the bellows tube, and this validity was proved in *in vivo* experiment.

Result of *in vivo* experiment indicated usability of our device in practical clinical condition. The outer sheath could go into gaps between organs, abdominal wall and membranes in deep cavity with wire-driven bending motion of the distal end as switching the flexible and the rigid modes, and flexible instruments such as fiberscope and biopsy forceps could be passed through the bellows tube into target deep space. On the other hand in manipulation of biopsy forceps, low resolution of the 2.8-mm fiberscope was not enough to observe local operative fields. And some methods of real-time secure detection of position and shape of the device were thought to be needed besides a normal laparoscope. In the outer sheath, some small magnetic positioning sensors are mountable to track approach path of the device. Or more specifically, MRI-guided surgery is possible due to MRI-compatibility of the device.

Moreover our device is useful for natural orifice transluminal endoscopic surgery (NOTES) with improvements to downsize the outer diameter under 10 mm and to dilate inner channel. In NOTES, transluminal approach path is generally twisted up in 3D space and it is significant to transport required flexible instruments to deep operative field in safely. Our device can provide an arbitrary secure path for various flexible instruments by itself, and size of inner channel diameter, total length and length of active-bending distal end are customized freely for on demand minimally invasive surgery.

5 Conclusion

In order to secure an approaching path for minimally invasive surgical instruments especially used in deep operative fields, we developed a nonmagnetic rigid and flexible outer sheath with pneumatic interlocking mechanism using flexible toothed links and an active-bending wire-driven distal end. We evaluated mechanical performances of bending angle and holding power of the device to use in practical clinical condition. In vivo

experiment using a swine indicated validity of large bending angle up to ±180 degrees and sufficient holding power up to 3.6 N in the flexible and the rigid modes, and that some flexible instruments can be inserted into deep areas in abdominal cavity. These results show sufficient possibility of use of the device in NOTES with some improvements in tracking method for minimally invasive image-guided surgery.

Acknowledgment. We wish to thank Dr. Hashizume, Dr. Tomikawa, Dr. Tanoue, Dr. Ieiri and Dr. Konishi in Kyushu University for cooperation in vivo experiment. This work was supported in part by Grant-in-aid for Scientific Research (18680042, 1710008) of Ministry of Education, Culture, Sports, Science and Technology (MEXT) in Japan, the Special Coordination Funds for Promoting Science and Technology commissioned by the MEXT in Japan, Grant-in-Aid for Scientific Research (20650077) of the MEXT in Japan, and Grant for Industrial Technology Research (07C46050), New Energy and Industrial Technology Development Organization, Japan.

References

1. Peirs, J., Reynaerts, D., Van Brussel, H.: A Miniature Manipultor for Integration in a Self-propelling Endoscope. Sensors and Actuators A 92, 343–349 (2001)
2. Nakamura, R., Kobayashi, E., Masamune, K., Sakuma, I., Dohi, T., Yahagi, N., Tsuji, T., Hashimoto, D., Shimada, M., Hashizume, M.: Multi-DOF forceps manipulator system for laparoscopic surgery. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 653–660. Springer, Heidelberg (2000)
3. Yamashita, H., Hata, N., Kim, D., et al.: Handheld Laparoscopic Forceps Manipulator Using Multi-slider Linkage Mechanisms. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 121–128. Springer, Heidelberg (2004)
4. Abbou, C.C., Hoznek, A., Salomon, L., et al.: Laparoscopic radical prostatectomy with a remote controlled robot. Journal of Urology 165, 1964–1966 (2001)
5. Ikuta, K., Sasaki, K., Yamamoto, K., et al.: Remote Microsurgery System for Deep and Narrow Space - Development of New Surgical Procedure and Microrobotic Tool. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 163–172. Springer, Heidelberg (2002)
6. Nakamura, Y., Matsui, A., Saito, T.: Shape Memory-Alloy Active Forceps for Laparoscopic Surgery. In: Proc. 1995 IEEE International Conference on Robot and Automation, vol. 3, pp. 2320–2327 (1995)
7. Simaan, N., Taylor, R., Flint, P.: High Dexterity Snake-Like Robotic Slaves for Minimally Invasive Telesurgery of the Upper Airway. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 17–24. Springer, Heidelberg (2004)
8. Sturges, R.H., Laowattana Jr., S.: A flexible, tendon-controlled device for endoscopy. In: Proc. 1991 IEEE International Conference on Robotics and Automation, vol. 3, pp. 2582–2591 (1991)
9. Amir, D., Howie, C., Alon, W., et al.: Percutaneous intrapericardial interventions using a highly articulated robotic probe. In: The First IEEE/RAS-EMBS International Conference on Biomedical Robotics and Biomechatronics, pp. 7–12 (2006)
10. Yagi, A., Matsumiya, K., Masamune, K., et al.: Rigid-Flexible Outer Sheath Model Using Slider Linkage Locking Mechanism and Air Pressure for Endoscopic Surgery. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 503–510. Springer, Heidelberg (2006)
11. Hata, N., Tokuda, J., Hurwitz, S., et al.: MRI-Compatible Manipulator with Remote-Center-of-Motion Control. J. Magn. Reson. Imaging 27(5), 1130–1138 (2008)
12. De Visser, H., Heijnsdijk, E.A.M., Herder, J.L., et al.: Forces and displacements in colon surgery. Surgical Endoscopy 16, 1426–1430 (2002)

Data-Derived Models for Segmentation with Application to Surgical Assessment and Training

Balakrishnan Varadarajan[1], Carol Reiley[2], Henry Lin[2], Sanjeev Khudanpur[1,2], and Gregory Hager[2]

[1] Department of Electrical and Computer Engineering
[2] Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA
{bvarada2,creiley,hcl,khudanpur,hager}@jhu.edu

Abstract. This paper addresses automatic skill assessment in robotic minimally invasive surgery. Hidden Markov models (HMMs) are developed for individual surgical gestures (or *surgemes*) that comprise a typical bench-top surgical training task. It is known that such HMMs can be used to recognize and segment surgemes in previously unseen trials [1]. Here, the topology of each surgeme HMM is designed in a data-driven manner, mixing trials from multiple surgeons with varying skill levels, resulting in HMM states that model skill-specific *sub-gestures*. The sequence of HMM states visited while performing a surgeme are therefore indicative of the surgeon's skill level. This expectation is confirmed by the average edit distance between the state-level "transcripts" of the same surgeme performed by two surgeons with different expertise levels. Some surgemes are further shown to be more indicative of skill than others.

1 Automatic Skill Assessment in Robotic Surgery

Robotic minimally invasive surgery (RMIS) has experienced rapid development and growth over the past decade, and the da Vinci robotic surgery system has emerged as the leader in RMIS [2]. Training for RMIS has often been cited as challenging, even for experienced surgeons [3]. One approach to overcome this challenge is develop techniques for automatic assessment of surgical skills during the performance of benchmark tasks, such as suturing or knot-tying, that simulate live tasks used for clinical skill evaluation [4]. This paper presents such techniques based on gesture recognition using hidden Markov models (HMMs).

RMIS is uniquely amenable to automatic skill assessment. The robot functions as a *measurement tool* for dexterous motion. As part of its run-time system, the da Vinci exposes an application programming interface (API) which provides accurate and detailed kinematic motion measurements, including the surgeon console "master" manipulators and all patient-side tools. We use these measurements to recognize individual surgical gestures [1]. Using both surgeon- and patient-side kinematics may seem redundant. But since one may carry some information that the other doesn't, (e.g intended v/s actual tool motion), we use both, and apply data-driven dimensionality reduction techniques to remove such redundancies.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 426–434, 2009.

Dosis et al [5] have used hidden Markov models to model hand manipulations and to classify simple surgical tasks. Richards et al [6] have demonstrated that force/torque signatures may be used in RMIS for two-way skill classification. Rosen et al [7] have used HMMs to model tool-tissue interactions in laparoscopic surgery; a seperate HMM for each skill level was trained using a pool of surgeons, and a *statistical distance* between these HMMs was shown to correlate well with the learning curve of these trainee surgeons. In these and other reported efforts, the automatic assessment is for entire trials, while the work presented here assesses *finer grained* segments, namely individual surgical gestures.

Lin et al [8] have used linear discriminant analysis (LDA) to project the high-dimensional kinematic measurements from the da Vinci API to three or four dimensions, and used a Bayes' classifier to segment surgical gestures from the low-dimensional signal. Reiley et al [1] replace their Bayes classifier with a 3-state left-to-right HMM for each gesture, and demonstrate improved accuracy on unseen users. The work presented here improves upon [1] by performing LDA to discriminate between the kinematical signal of *sub-gestures* – modeled by individual HMM states – rather than between the signal of entire gestures.

The distinguishing contribution of this work is the application of the HMM methodology to gesture-specific skill assessment. A data-driven algorithm is used to design the HMM topology for each gesture. As a consequence, in addition to automatic detection and segmentation of surgical gestures, one is able to compare individual gestures of expert, intermediate and novice surgeons in a quantitative manner. For instance, some gestures in a suturing task, such as navigating a needle through the tissue, are demonstrated to be more indicative of expertise than others, such as pulling the thread. Such fine grained assessment can ultimately lead to better automatic surgical assessment and training methods.

This paper is organized as follows. We begin in Section 2 with a background review of the suturing task and the use of HMMs for gesture recognition and segmentation. We then describe the two technical novelties in the use of HMMs, namely state-specific LDA and data-derived HMM topologies, in Section 3. This leads to improved gesture recognition accuracies. In Section 4, we demonstrate how paths through the HMM state space are indicative of the expertise with which the gesture has been performed, leading to the main contribution of the paper: a framework for automatic, gesture-level surgical skill assessment.

2 Surgical Gesture Recognition Using HMMs

2.1 The Surgeme Recognition Experimental Setup

Kinematic Data Recordings: We recorded the kinematic measurements from 2 expert, 3 intermediate and 3 novice surgeons performing a bench-top suturing task—four stitches along a line—on the teleoperated da Vinci surgical system. The average duration of a trial is 2 minutes, and the video and kinematic data are recorded at 30 frames per second. The kinematic measurements include position, velocity, etc. from both the surgeon- and patient-side manipulators for a total of 78 motion variables. We use $\{y_t, t = 1, 2, \ldots, T\}$ to denote the sequence of

kinematic measurements for a trial, with $y_t \in \mathbb{R}^{78}$ and $T \approx 3400$. A total of 30 trials were recorded, roughly four from each of the eight surgeons.

Manual Labeling of Surgemes: Each trial was manually segmented into semantically "atomic" gestures, based on the eleven-symbol vocabulary proposed by [1]. Following their terminology, we will call each gesture a *surgeme*. Typical surgemes include, for instance, (i) positioning the needle for insertion with the right hand, (ii) inserting the needle through the tissue till it comes out where desired, (iii) reaching for the needle-tip with the left hand, (iv) pulling the suture with the left hand, etc. We use $\{\sigma_{[i]}, i = 1, 2, \ldots, k\}$ to denote the surgeme label-sequence of a trial, with $\sigma_{[i]} \in \{1, \ldots, 11\}$ and $k \approx 20$, and $[b_i, e_i]$ the begin- and end-time of $\sigma_{[i]}$, $1 \leq b_i < e_i \leq T$. Note that $b_1 = 1$, $b_{i+1} = e_i + 1$, $e_k = T$.

The Surgeme Recognition Task: Given a partition of the 30 trials into *training* and *test* trials, the surgeme recognition task is to automatically assign to each trial in the test partition a surgeme transcript $\{\hat{\sigma}_{[i]}, i = 1, 2, \ldots, \hat{k}\}$ and time-marks $[\hat{b}_i, \hat{e}_i]$. Trials in the training partition are used to train the HMMs, as described below. We report results with three different training/test partitions.

Setup I: Of the 30 trials, 8 have some minor errors by the surgeons during suturing. These are excluded altogether in Setup I. Leave-one-out cross-validation is carried out with the remaining 22 trials, so that each trial is once in the test partition. The test results of all 22 folds (22 trials) are aggregated.

Setup II: The training partition in Setup II comprises the 22 "good" trials, while the test partition comprises only the 8 "imperfect" trials.

Setup III: User-disjoint partitions of the 30 trials are created in Setup III. An eight-fold cross validation akin to Setup I is carried out, except that in each fold, all the trials of 1 surgeon are in the test partition and all trials of the remaining 7 surgeons are in training. Test results of all 30 trials are aggregated.

Setup I is relatively the easiest, with 22 good test trials and the surgeon of each test trial seen in training. Setup II is harder, with seen surgeons but with test trials that have some visible errors, a situation not dissimilar from recognition of slightly disfluent speech. Setup II is most similar to the multiple-user results in [1, Table 3], with which we make direct comparisons. Setup III is the hardest, because all trials of the test surgeon have also been removed from training.

Recognition accuracy is measured as the fraction of kinematic frames that are assigned the correct surgeme label by an automatic system. Formally,

$$\text{Accuracy of test trial } \{y_1, \ldots, y_T\} = \frac{1}{T} \sum_{t=1}^{T} \mathbb{I}\left(\sigma_t = \hat{\sigma}_t\right), \tag{1}$$

where $\sigma_t = \sigma_{[i]}$ for all $t \in [b_i, e_i]$ and $\hat{\sigma}_t = \hat{\sigma}_{[i]}$ for all $t \in [\hat{b}_i, \hat{e}_i]$. This measures the goodness of both the labels *and* the segmentation proposed by $\{\hat{\sigma}_t\}$.

2.2 HMM-Based Surgeme Recognition

Dimensionality Reduction: Before surgeme recognition, the 78-dimensional kinematic data are reduced to $d \ll 78$ dimensions via LDA [9]. Specifically, each

block of $2p+1$ frames in the training partition is converted into a data-label pair $\left([y_{t-p}^{\mathrm{T}} \ldots y_{t-1}^{\mathrm{T}}\, y_t^{\mathrm{T}}\, y_{t+1}^{\mathrm{T}} \ldots y_{t+p}^{\mathrm{T}}]^{\mathrm{T}}, \sigma_t\right)$, and a $d \times 78(2p+1)$ projection matrix $\mathbf{A}$ is computed that maximizes the ratio of between- and within-surgeme scatter of the *projected* data $x_t = \mathbf{A}[y_{t-p}^{\mathrm{T}} \ldots y_t^{\mathrm{T}} \ldots y_{t+p}^{\mathrm{T}}]^{\mathrm{T}}$. Typically, $p = 5$ and d is 3 to 10. The $\{x_t\}$ are used everywhere subsequently, instead of $\{y_t\}$.

Surgeme Modeling: The likelihood of the kinematic signal $\{x_t, t = b_i, \ldots, e_i\}$ of a surgeme $\sigma_{[i]} = \sigma$ is modeled via a HMM as

$$P_\sigma(x_{b_i}, \ldots, x_{e_i}) = \sum_{s_{b_i} \in \mathcal{S}_\sigma} \sum_{s_{b_i+1} \in \mathcal{S}_\sigma} \cdots \sum_{s_{e_i} \in \mathcal{S}_\sigma} \prod_{t=b_i}^{e_i} p(s_t|s_{t-1}) \mathcal{N}(x_t\,;\, \mu_{s_t}, \Sigma_{s_t}), \quad (2)$$

where $\mathcal{S}_\sigma$ denotes the *hidden states* of the model for surgeme σ, $p(s|s')$ are the transition probabilities between these states, and $\mathcal{N}(\cdot\,;\, \mu_s, \Sigma_s)$ is a multivariate Gaussian density with mean μ_s and covariance Σ_s associated with state $s \in \mathcal{S}_\sigma$.

Parameter Estimation: Kinematic data from all training samples of a surgeme σ are modeled by the same HMM (with states $\mathcal{S}_\sigma$), and each surgeme is modeled by a different HMM. Model parameters are chosen to maximize the likelihood (2) of the training data $\{x_t\}$ via the standard Baum-Welch algorithm [10].

Surgeme Recognition: A surgeme (HMM) is permitted to be followed by any other surgeme during recognition, and the Viterbi algorithm [10] is used to find the sequence $\{\hat{s}_t \in \bigcup \mathcal{S}_\sigma, t = 1, \ldots, T\}$ of HMM states with the highest *a posteriori* likelihood given a test trial $\{x_t\}$. The surgeme sequence $\{\hat{\sigma}_{[i]}, i = 1, 2, \ldots, \hat{k}\}$ and time-marks $[\hat{b}_i, \hat{e}_i]$ are a byproduct of the Viterbi algorithm.

3 Improved Dimensionality Reduction and Modeling

3.1 Linear Discriminant Analysis Based on HMM States

The primary purpose of LDA is to reduce the dimensionality of $\{y_t\}$ without losing information necessary to discriminate between gestures σ_t. Note, however, that each surgeme is modeled by a HMM with several states $s \in \mathcal{S}_\sigma$, each of which models a sub-gesture—called a *dexeme* to connote small dextrous motions. It is natural, therefore, to investigate whether it is better to perform LDA to discriminate between dexemes rather than entire surgemes. An immediate hurdle we face is that the manual segmentation of $\{y_t\}$ is only up to surgemes, and not at the finer resolution of dexemes. But the HMM formalism provides a workaround.

Using the d-dimensional training data $\{x_t\}$ derived from surgeme-level LDA, we first estimate surgeme HMMs as described above, and use the Viterbi algorithm to obtain a *forced alignment* of $\{x_t\}$ with the states of the surgeme HMMs. This results in a dexeme-level segmentation of each surgeme. We use the resulting dexeme *label* $\hat{s}_t$ of each block $[y_{t-p}^{\mathrm{T}} \ldots y_{t-1}^{\mathrm{T}}\, y_t^{\mathrm{T}}\, y_{t+1}^{\mathrm{T}} \ldots y_{t+p}^{\mathrm{T}}]^{\mathrm{T}}$ to compute a new projection matrix $\mathbf{A}$ and use that for all subsequent experiments.

The dexeme-level LDA is better able to preserve information that distinguishes temporal sub-gestures of a single gesture, as well as stylistic variations between samples of the same gesture, as will be demonstrated in Section 3.3.

3.2 Data-Derived HMM Topologies

In the work of [1], and in our initial work here, we used a 3-state left-to-right HMM to model each gesture. However, each gesture has not only temporally distinct sub-gestures—which would be well modeled by states of a left-to-right HMM—but also *contextual* variability in sub-gestures. Some of the latter variability is due to the skill level of the surgeon, some due to the dynamics of a previous or subsequent gesture, while some depends on where in the suturing task (e.g. on the first or fourth stitch) the gesture is being performed. We investigate induction of an optimal HMM topology directly from the data to model such variability.

Formally, we wish to find the topology of a surgeme HMM that maximizes the likelihood (2) of the training data $\{x_t\}$. Finding the optimal HMM topology, however, is computationally intractable: given $n = |\mathcal{S}_\sigma|$, one must find, separately for every n-vertex directed graph, the HMM parameters that maximize (2).

In Speech recognition, HMM topologies are derived for capturing context-dependent (allophonic) variations of phonemes using greedy algorithms. We apply one such algorithm by Varadarajan et al [11], called the modified successive state splitting (SSS) algorithm, to our problem. We begin with a single-state HMM for each surgeme, and iteratively estimate the HMM parameters and increment the number of HMM states via SSS .

Data-derived HMM topologies yield accurate models for surgeme recognition, and also capture sub-gesture patterns indicative of skill, as shown in Section 4.

3.3 Surgeme Recognition and Segmentation Results

We performed surgeme recognition experiments with the training/test partitions described in Section 2.

We first estimated a 1-state HMM per surgeme. In this case, there is no difference between surgeme-level and dexeme-level LDA. The 70% to 74% accuracy for Setup II reported in Table 1(a) may therefore be directly compared with the results of [1], who report accuracies of 64% to 72%.

Next, we estimated a 3-state left-to-right HMM for each surgeme. With surgeme-level LDA, [1] report accuracies of 72% to 77%. In comparison, the dexeme-level LDA provides up to 86% accuracy, as shown in Table 1(b). We also see from Table 1(b) that maximum accuracy is achieved when the number of dimensions d is between 9 and 17 indicating the need for more dimensions to differentiate between the finer grained motions represented by dexemes.

Modeling a surgeme as a temporal sequence of 3 dexemes (left-to-right HMM states) is better than a single-state HMM, but still ad hoc. Determining the HMM topology from data permits modeling both temporally distinct sub-gestures and contextual variability of gestures, as discussed in Section 3.2. Therefore, we use

Table 1. Surgeme Recognition Accuracies with Dexeme-level LDA

(a) A 1-state HMM per Surgeme

LDA d	Setup I	Setup II	Setup III
3	75%	75%	58%
5	81%	72%	69%
7	81%	70%	72%

(b) A 3-state HMM per Surgeme

LDA d	Setup I	Setup II	Setup III
3	79%	70%	73%
5	82%	76%	73%
7	82%	83%	81%
9	82%	**86%**	78%
17	**87%**	83%	**81%**

(c) Data-derived HMM Topology

LDA d	Setup I	Setup II	Setup III
3	69%	67%	64%
4	73%	73%	70%
10	83%	82%	**73%**
15	86%	82%	71%
20	**87%**	**83%**	70%

the SSS algorithm to evolve a 6-state HMM for each gesture. Table 1(c) shows recognition results for the different setups. The recognition accuracies remain high for Setup I and II using data-derived HMMs. The maximum recognition accuracy is obtained when the number of dimensions d is 20, indicating the need for more dimensions needed to differentiate between the larger number of dexemes. We also note that the accuracies drop considerably for Setup III. We conjecture that in addition to expertise-dependent dexemes, the data-derived HMMs may also be modeling user-specific dexemes. This leads to improved recognition when a new trial of a seen user is presented, but also to some *overfitting* to seen users.

The optimal LDA dimension is empirically seen to be proportional to the number of classes: 5 for 1-state HMMs (discriminating 8 surgemes), 9-17 for 3-state HMMs (24 dexemes), and 15-20 for data-derived HMMs (48 dexemes).

4 Surgeme-Level Skills Revealed in Dexeme-Sequences

To illustrate how data-derived HMM topologies encode dexterity information, consider Figure 1, which shows a 5-state HMM derived via the SSS algorithm for surgeme #3 corresponding to the act of "inserting needle through the tissue.". Training samples of surgeme #3 were aligned with this 5-state HMM, and the state-level time marks were used to isolate individual dexemes corresponding to the HMM states a, b, c, d and $e \in \mathcal{S}_3$.

We studied the endoscope video to understand what the segments that align with each dexeme (HMM state) represent, and observed the following.[1]

Dexemes a, b and c: They all constitutes rotating of the right hand patient-side wrist to drive the needle from the entry- to the exit.

Dexeme c versus a and b: All examples that aligned to c were from *novice* surgeons. Examining the videos revealed that c corresponds to a sub-gesture where the novice hesitates/retracts while pushing the needle to the exit point. In most cases, c is followed by a or b, in which the trainee surgeon eventually performs

[1] Video corresponding to these dexemes is available at www.clsp.jhu.edu/~balakris/MICCAI2009/

the task (inserting the needle till it exits) correctly. States a and b appear to be indistinguishable, except for some stylistic differences.

Dexeme d: It represents the left arm reaching for the exiting needle. Often, when the left arm is already positioned near the exit point, this gesture is omitted. This explains the transitions from states a and b directly to state e.

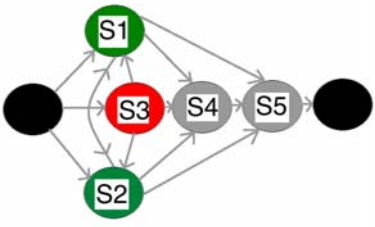


Fig. 1. The Data-derived HMM for $n = 5$ States for Gesture #3

Dexeme e: It represents firmly gripping the needle with the left arm.

These observations reinforce the claim that *SSS provides a means for automatically inducing meaningful units for modeling dexterous motion*. While not demonstrated here, it may be applied to entire trials, automatically discovering and modeling gestures without requiring any manual labeling!

4.1 Measuring Expertise by Aligning Dexeme-Transcripts

To compare how dissimilar two instances of a surgeme are, we compute an edit distance between their *dexeme transcripts* as described below.

Let $\{x_t^1, t = \hat{b}_i, \ldots, \hat{e}_i\}$ and $\{x_t^2, t = \hat{b}_j, \ldots, \hat{e}_j\}$ denote two *automatically* segmented and labeled realizations of the surgeme σ, i.e. $\hat{\sigma}_{[i]} = \hat{\sigma}_{[j]} = \sigma$. We use the Viterbi alignment of $\{x_t^1\}$ with the states $\mathcal{S}_\sigma$ of the surgeme HMM to obtain the sequence $\{\hat{s}_t^1, t = \hat{b}_i, \ldots, \hat{e}_i\}$, and similarly $\{\hat{s}_t^2, t = \hat{b}_j, \ldots, \hat{e}_j\}$ from $\{x_t^2\}$. We then obtain the sequence of HMM states visited by $\{x_t^1\}$ (resp. $\{x_t^2\}$) by simply compacting each *run* of state labels. In other words, we ignore how many *consecutive* frames are aligned with a state, counting them collectively as one "visit" to the state. Let $\{\hat{s}_{[i]}^1, i = 1, \ldots, \hat{k}^1\}$ and $\{\hat{s}_{[j]}^2, j = 1, \ldots, \hat{k}^2\}$ denote the dexeme transcripts of the two gestures generated in this manner.

We then align $\{\hat{s}_{[i]}^1\}$ and $\{\hat{s}_{[j]}^2\}$ using Levenshtein distance, and each element in the two sequences is marked as matched if it is aligned with the an identical element in the other sequence. Inserted, deleted and (both sides of a pair of) mismatched symbols are marked as mismatched. The *similarity* of the realizations $\hat{\sigma}_{[i]}$ and $\hat{\sigma}_{[j]}$ is defined as the number of matched dexemes divided by $\hat{k}^1 + \hat{k}^2$. A similarity of 1 corresponds to identical dexeme sequences: $\hat{k}^1 = \hat{k}^2$ and $\hat{s}_{[i]}^1 = \hat{s}_{[i]}^2$ for each i. Otherwise similarity ranges between 0 and 1.

We calculate the average edit distance between realizations of σ drawn from different expertise levels for the four most frequent gestures: $\sigma = 2, 3, 4$ and 6.

Note from Tables 2(a), 2(b) and 2(c) that some surgemes (e.g. #2 : "positioning the needle at the entry point" or #3 : "inserting the needle through the tissue") show low expert-novice similarity compared to expert-expert, indicating the need for skillful execution. In comparison, surgeme #6 (pulling the suture) in Table 2(d) exhibits significant similarity even between experts and novices.

The correlation between expertise level and edit distance is clearly evident.

Table 2. Dexeme Similarity of Surgemes Performed with Different Skill Levels

(a) Similarities in Surgeme #2

	Expert	Inter.	Novice
Expert	0.65	0.55	0.55
Intermediate	0.55	0.50	0.53
Novice	0.55	0.53	0.46

(b) Similarities in Surgeme #3

	Expert	Inter.	Novice
Expert	0.69	0.60	0.53
Intermediate	0.60	0.51	0.50
Novice	0.53	0.50	0.50

(c) Similarities in Surgeme #4

	Expert	Inter.	Novice
Expert	0.71	0.57	0.54
Intermediate	0.57	0.58	0.58
Novice	0.54	0.58	0.51

(d) Similarities in Surgeme #6

	Expert	Inter.	Novice
Expert	0.74	0.69	0.68
Intermediate	0.69	0.65	0.67
Novice	0.68	0.67	0.61

5 Concluding Remarks and Potential Applications

We have demonstrated the utility of sub-gesture-level LDA in improving dimensionality reduction for HMM-based gesture recognition. We have also shown that data-derived HMMs automatically discover and model skill-specific sub-gestures, leading to a natural metric (dexeme edit distance) for comparing surgical gestures for skill assessment. Since the dexemes are data-derived, such comparison may be feasible even if the manual labeling of surgemes is very coarse grained or absent. Finally, dexeme edit distance based alignment may be transferred to synchronize the surgical *video*, opening up immense possibilities for training.

References

1. Reiley, C., Lin, H., Varadarajan, B., Khudanpur, S., Yuh, D.D., Hager, G.D.: Automatic recognition of surgical motions using statistical modeling for capturing variability. In: MMVR (2008)
2. Shuford, M.: Robotically assisted laparoscopic radical prostatectomy: a brief review of outcomes. Proc. Baylor University Medical Center 20(4), 354–356 (2007)
3. Lenihan Jr., J., Kovanda, C., Seshadri-Kreaden, U.: What is the Learning Curve for Robotic Assisted Gynecologic Surgery? J. Min. Inv. Gyn. 15(5), 589–594 (2008)
4. Martin, J., Regehr, G., Reznick, R., MacRae, H., Murnaghan, J., Hutchison, C., Brown, M.: Objective structured assessment of technical skill (OSATS) for surgical residents. British Journal of Surgery 84(2), 273–278 (1997)
5. Dosis, A., Bello, F., Gillies, D., Undre, S., Aggarwal, R., Darzi, A.: Laparoscopic task recognition using hidden markov models. In: MMVR (2005)
6. Richards, C., Rosen, J., Hannaford, B., Pellegrini, C., Sinanan, M.: Skills evaluation in minimally invasive surgery using force/torque signatures. Surgical Endoscopy 14, 791–798 (2000)
7. Rosen, J., Solazzo, M., Hannaford, B., Sinanan, M.N.: Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden markov model. Computer Aided Surgery 7(1), 49–61 (2002)

8. Lin, H.C., Shafran, I., Murphy, T.E., Okamura, A.M., Yuh, D.D., Hager, G.D.: Automatic detection and segmentation of robot-assisted surgical motions. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 802–810. Springer, Heidelberg (2005)
9. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals of Eugenics 7, 179–188 (1936)
10. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257–286 (1989)
11. Varadarajan, B., Khudanpur, S., Dupoux, E.: Unsupervised learning of acoustic sub-word units. In: Proceedings of ACL 2008: HLT, Short Papers, 165–168 (2008)

Task versus Subtask Surgical Skill Evaluation of Robotic Minimally Invasive Surgery

Carol E. Reiley and Gregory D. Hager

Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218, USA
{creiley,hager}@jhu.edu

Abstract. Evaluating surgical skill is a time consuming, subjective, and difficult process. This paper compares two methods of identifying the skill level of a subject given motion data from a benchtop surgical task. In the first method, we build discrete Hidden Markov Models at the *task level*, and test against these models. In the second method, we build discrete Hidden Markov Models of surgical gestures, called *surgemes*, and evaluate skill at this level. We apply these techniques to 57 data sets collected from the da Vinci surgical system. Our current techniques have achieved accuracy levels of 100% using task level models and known gesture segmentation, 95% with task level models and unknown gesture segmentation, and 100% with the surgeme level models in correctly identifying the skill level. We observe that, although less accurate, the second method requires less prior label information. Also, the surgeme level classification provided more insights into what subjects did well, and what they did poorly.

1 Introduction

Human motion is stochastic in nature. A person performing a repeatable task multiple times (e.g. drawing a straight line) would generate different motion measurements (ie. forces, velocities, positions, etc.) despite the fact that the measurements represent the the same task performed with the same level of skill. The goal of skill modeling is to uncover and measure the underlying characteristics of skill hidden in measurable motion data. In this paper, we focus on modeling and assessing surgical technical skill. Current techniques for surgical skill assessment include descriptive statistics (time, path length, number of motions), morbidity rates, and checklists [1,2]. However, these methods require manual interpretation, lack flexibility, and are time consuming and labor intensive. We note that, robotic surgery, in particular, is known to have a steep learning curve and be difficult to teach [3]. However, robotic surgery is reported to be the fastest growing segment of computer aided surgical systems - an industry expected to grow to $2 billion by the year 2010 [4]. Thus, automated assessment and training will have a potentially high impact in this area.

In what follows, we make extensive use of Hidden Markov Models (HMMs) for statistical modeling of time-series motion data as a basis for skill assessment. HMMs are statistical models used to determine hidden parameters from observed data. An HMM can either be continuous or discrete. A discrete HMM, which is used in this work, is represented by $\lambda = (A, B, \Pi)$ where A is the state transition probability distribution matrix, B is the observation symbol probability distribution matrix, and Π is the initial

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 435–442, 2009.

state distribution [5]. They are extensively developed in the area of speech recognition [5] and have proved useful in studying teleoperation and human skill evaluation for non-surgical tasks [6,7,8,9]. Recently, HMMs have been used in laparoscopic [10] and virtual simulators [11,12] to classify the skill of a surgeon. Motivated by these studies, we propose building HMMs driven by expertise examples on teleoperated robotic systems.

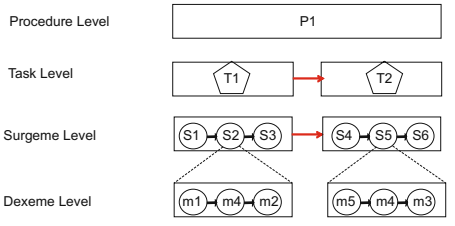


Fig. 1. A hierarchical decomposition of surgical motion. Modeling can be done on tasks (e.g. suturing), their decomposition into surgemes (e.g. needle pulling), and even more primitive motion elements called dexemes.

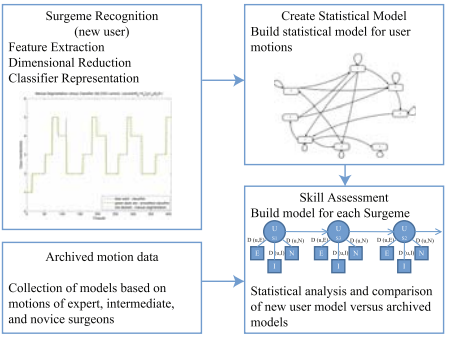


Fig. 2. Flowchart of building skill evaluation models

In prior work [13,14], it has been proposed that surgical tasks can be broken down into a hierarchy of more primitive gestures, sometimes referred to as "surgemes" (Figure 1). To the best of our knowledge, this is the first work that has analyzed surgical models on the surgeme level. Thus, we are evaluating if skill at the task level or at the surgeme level provides more accuracy or information. At either level, we may also explore whether knowing the underlying surgeme information at training and/or testing [10] provides significant additional information. Figure 2 is a general block diagram of building skill models.

In this paper, we develop skill-dependent HMM models for three levels of surgical expertise, novice, intermediate, and expert. We define an expert as a practicing surgeon with more than 100 hours of clinical surgical robotic experience. An intermediate as either a fellow or resident surgeon with less than 100 hours of surgical robotic experience, and a novice as a non-surgeon with no prior robotic experience. As shown in Table 1 this leads to six possible approaches to surgical skill evaluation. We then explore the relationships between three of these problem settings:

Table 1. Possible combinations of labeling information and level of skill evaluation

Training Testing	Labeled Labeled	Labeled Unlabeled	Unlabeled Unlabeled
Surgeme Level	1a	1b	1c
Task Level	2a	2b	2c

Experiment 1: Surgeme-Level Hidden Markov Models (1c - states: unknown dexemes) vs. Task-Level Hidden Markov Models (2c - states: unknown surgemes). We hypothesize that modeling skill on the surgeme level may provide insight to what portions of a task a subject performs proficiently or where he/she performs like a novice.

Experiment 2: Task-Level Hidden Markov Models With Known States (2a - states: motion surgemes defined in Experimental Setup) vs. Task-Level Hidden Markov Models With Unknown States (2c). Here, we investigate whether raw motion data can be modeled and evaluated for skill level *without* prior manual labeling.

2 Experimental Setup

Surgical Platform. The da Vinci surgical system by Intuitive Surgical Inc. [15] is a clinical, teleoperated robot used in operating rooms worldwide. It is a two handed manipulator with 7 degrees of freedom each. Due to its immersive interface, it provides a structured, well instrumented, unobtrusive environment for studying surgical motions. Using the da Vinci application programming interface (API), synchronized high-resolution video and motion data were recorded at 23 Hz and resampled to 40 Hz for data analysis. For the results reported below, we use a 14 variable subset of the available motion channels comprising of six joint velocity values and gripper information of the patient-side left and right robotic manipulators. Fifty seven trials of a four-throw suturing task were collected from a group of nine different surgeons categorized into three different expertise levels; 19 trials each for expert, intermediate, and novice. (Fig. 3).

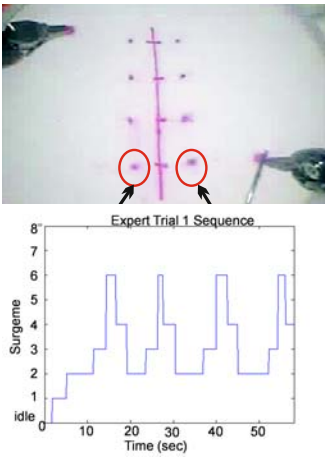


Fig. 3. (top) Experimental setup displaying surgeme 1 (reach for needle) in our motion vocabulary. (bottom) Example of manually labeled trial using surgemes

We make use of the motion vocabulary defined in [13] which consists of (0) idle position, (1) reach for needle, (2) position needle, (3) inset needle through tissue, (4) transferring needle from left to right, (5) moving to center with needle in gripper, (6) pulling suture with left hand, (7) pulling suture with right, and (8) orienting needles. Idle motion time at the start and end of the trial (motion 0) was not used for data analysis. In order to have ground truth for training and validation, our data set was manually segmented based on the above surgeme motions. Not all trials were required to use all surgemes in the motion vocabulary. Trials times varied between 45 to 130 seconds.

3 Task-Level Skill Modeling

The aim of this paper is to create an HMM $\lambda = (A, B, \Pi)$ that describes the surgical performance made by surgeons of various skill levels and to create a metric to evaluate surgical performance.

Data Preprocessing. The 14 vector continuous motion observations were postprocessed into a discrete alphabet using vector quantization techniques similar to [6]. We first apply a Short Time Fourier Transform (STFT) on each of the 14 velocity signals, $x(t)$, over a 400 ms window every 200 ms to filter the high frequency data.

$$STFT\{x(t)\} \equiv X(t,\omega) = \int_{-\infty}^{\infty} x(t')w(t'-t)e^{-j2\pi ft'}\,dt' \qquad (1)$$

where $w(d) = 1$ if $0 < d < 200$ms and 0 otherwise. The STFT was chosen because it is able to extract useful features from time series data and, when applied over a sliding window in the time domain, was able to retain the time localization of events from the original data. The amplitudes of the lower 4 STFT coefficients were then concatenated to form a new 56 (14 velocity channels by 4 STFT coefficients) dimensional feature vector. Then we used the K-means algorithm [16] to search for a small number of cluster centers within each of the 9 states. These cluster centers form a so-called *codebook*. We chose a 64 symbol codebook based on empirical testing with values of 32, 64, and 128.

HMM Task Models. From the discretized signals, we trained two families of Hidden Markov Model for each skill level (expert λ_{s1}, intermediate λ_{s2}, and novice λ_{s3}) using the Matlab statistics toolbox [17]. In one family, we trained "true" Hidden Markov Models (2c) for each skill level using the Baum-Welch algorithm and no prior training labels. We chose 9 states to match the number of surgeme labels, but we allowed the generalized Expectation Maximization algorithm to uncover the best structure of the hidden states. For Hidden Markov Models with states as surgemes (2a), we concatenated each observation with its surgeme label. Because the underlying states were known, we did not need to perform Baum-Welch, but rather could directly calculate the transition and emission tables.

Evaluation Method. Given an observed output sequence O_{test} and a skill model λ_s, maximum log likelihood, $\log P(O_s|\lambda_s)$ was used to identify skill as

$$\lambda^* = \arg\max[\log P(O_{test}|\lambda_{se}), (\log P(O_{test}|\lambda_{si}), (\log P(O_{test}|\lambda_{sn})] \qquad (2)$$

where λ_{se},λ_{si},λ_{sn} are the expert, intermediate, and novice models for each trial, respectively.

Following [5,10], we also define a relative distance between a skill model λ_s and a model trained from the testing data λ_{test} based on the sequence of test observations

$$D(\lambda_s, \lambda_{test}) = \frac{1}{T_{test}} \min(\xi(\lambda_i, \lambda_{test}), \xi(\lambda_e, \lambda_{test}), \xi(\lambda_n, \lambda_{test})) \qquad (3)$$

where $\xi(\lambda_s, \lambda_{test}) = \log P(O_{test}|\lambda_{test}) - \log P(O_{test}|\lambda_s)$ and T_{test} is the length of the observation sequence O_{test}. This equation defines how well model λ_s matches observations generated by λ_{test}, relative to how well model λ_{test} matches observations generated by itself. Calculating the distance between a test sequence and skill levels: expert, intermediate, and novice yields three values. It is easy to see that the HMM model with minimum distance is also that with maximum log likelihood.

4 Surgeme-Level Skill Modeling

Data Preprocessing. The continuous motion values were discretized using a K-means algorithm with 8 cluster centers. The data was segmented into smaller motion blocks using the known manual surgeme labels, yielding a large set of observation subsequences for each surgeme.

HMM Surgeme Models. Three HMM expertise models for each surgeme were computed, totaling 24 skill models. We used 8 states after comparable results running the system with three, eight, and fourteen states.

Evaluation Method. In order to classify a test sequence as a particular skill level, we used Equation 2, where now λ_s is the trained surgeme expertise models and O_{test} is the observations from each surgeme occurrence. Each surgeme occurrence would have a corresponding skill label.

Each test trial now has a skill label associated to each surgeme occurrence. Majority voting was used on the test sequence labels, L_t, to classify the skill of that trial. In the event of a tie between skill levels, the maximum of the surgeme log-likelihood average was chosen.

5 Results

5.1 Experiment 1: Surgeme-Level HMM versus Task-Level HMM

o test the accuracy of the surgeme-based classification method, we performed a leave-one-out cross validation. In each round, one occurrence of a surgeme was left out for testing while the remaining occurrences of that surgeme was used to train an HMMs for each class. Out of the 1011 total surgeme occurrences, experts were correctly classified 75% (233/311), intermediates 59% (206/347), and novices 76% (268/353). For each expertise and surgeme, we present the number of correctly identified skill labels over the total number of occurrences of that surgeme (Table 2). The diagonal shows the correctly labeled surgemes. Reading across the matrix rows indicates the frequency of correctly classifying the skill of a surgeme and the frequency of it misclassified at another skill level. All rows sum to 1.

Table 2. Surgeme-based result: "Confusion Matrix"

	Exp.	Int.	Nov.	count
Exp. S1	0.50	0.28	0.22	18
Int. S1	0.33	0.67	0	18
Nov. S1	0.31	0	0.69	16
Exp. S2	0.76	0.12	0.12	76
Int. S2	0.16	0.78	0.06	77
Nov. S2	0.16	0.07	0.78	76
Exp. S3	0.79	0.17	0.04	76
Int. S3	0.35	0.53	0.12	75
Nov. S3	0.34	0.12	0.54	74
Exp. S4	0.89	0.02	0.09	57
Int. S4	0.00	0.78	0.22	27
Nov. S4	0.03	0.14	0.83	59
Exp. S5	-	0.25	0.75	4
Int. S5	0.11	0.79	0.11	19
Nov. S5	0.05	0.21	0.74	19
Exp. S6	0.71	0.08	0.22	78
Inter. S6	0.04	0.77	0.19	74
Nov. S6	0.05	0.17	0.79	42
Exp. S7	-	-	-	0
Inter. S7	-	0.92	0.08	36
Nov. S7	-	0.07	0.93	46
Exp. S8	-	-	1.0	2
Int. S8	-	0.76	0.24	21
Nov. S8	-	0.10	0.90	21

The data indicate that: (1) certain surgemes, such as 2, 4, 6, are indicative of skill based on high classification rates across all three skill levels, (2) other surgemes, such as 1, are not indicative of skill because the discrimination between skill levels is low, and (3) there are surgemes that are infrequently used by skill groups, such as surgemes 5,7,8 for an expert, suggesting that those are intermediate positioning moves. These results correlate to the distribution of time spent in each surgeme according to expertise. We found that regardless of the expertise level, most of the time was spent in surgeme 2 (Reaching for needle), then 3 (positioning needle), then 6 (Moving to center with needle in gripper). Incidentally, we also found that these surgemes were the ones that had the highest number of surgeme occurrences.

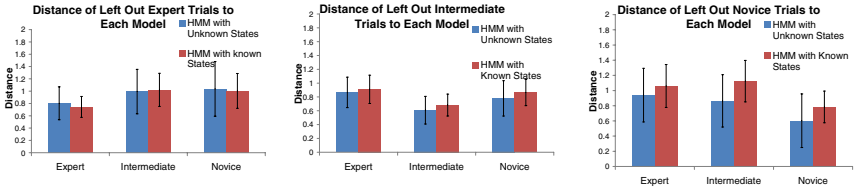


Fig. 4. Average distance, over 19 rounds, of the left out sequence to each class' model built using BW and KS

Table 3. Skill classification rates for trials using cross validation

Expertise	Classification Rate
1c: Surgeme (E)	100%
1c: Surgeme (I)	100%
1c: Surgeme (N)	100%
2c: Task BW(E)	84%
2c: Task BW (I)	100%
2c: Task BW(N)	100%
2a: Task KS(E)	100%
2a: Task KS (I)	100%
2a: Task KS(N)	100%

Intermediate skill was most difficult to classify correctly on the surgeme occurrence level as it had the highest percentage of misclassification overall, indicating that skill might be more accurately modeled as a scale instead of discrete classes. Assessing the surgemes in the context of their trials, we found that this evaluation system correctly labeled 100% of the trials. We found a 95% classification accuracy rate of the task-level with unknown states (BW) through 19 rounds of leave one out. In each round one trial from each class was left out. HMMs were built for each expertise and each left one trial out testing. Three expert trials were misclassified as a novice.

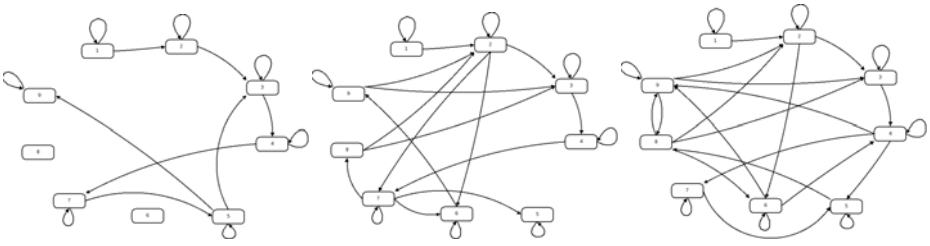
5.2 Experiment 2: Task-Level Hidden Markov Model with Known (KS) versus Unknown States (BW)

We performed 19 rounds of leave-one-out testing to determine the accuracy of this classification method. In each round, one trial from each class was left out. HMMs were built for each expertise and each left out test trial was tested using KS methods compared to BW.

Even evaluating the state transition diagrams, we see a qualitative difference between the movements of experts and novices (Figure 5). For the KS transition diagram, as surgeon skill increases; the graph of their movements becomes more directed. The expert surgeon accomplished a task using relatively few movements whereas the novice made more errors during the task causing them use extraneous motions and started over in the initial states.

There is still a difference between the expert and novice models using the Baum-Welch algorithm where the states are hidden. Even though the states do not correspond to the labeled surgemes it can still be observed that the expert model is less connected than the novice model. This shows that the movements of experts are more directed and less prone to erroneous or unnecessary movements.

Figure 4 shows the average distance over the 19 rounds of the expert, intermediate and novice trials being generated by each expertise model. The first of the three sets of



(a) Exp.KS (b) Int.KS (c) Nov.KS

Fig. 5. Expertise state transition diagram with known surgemes

bars represents the BW algorithm, and the second the KS algorithm. Each skill level had the minimum distance when compared against its own model. The intermediate skill level surgeon has a smaller distance to the expert than the novice surgeon. This seems to confirm that intermediate surgeons actually do perform the tasks in a manner that is closer to the expert than the novice does. KS yielded 100% trial classification accuracy which is an improvement over the 95% classification using BW.

6 Discussion and Conclusions

This paper addresses the problem of evaluating skill using continuous velocity data from the da Vinci system. Vector quantization techniques were used to discretize the data to train discrete Hidden Markov Models (HMMs). These experiments show that HMMs are a useful method to classify skill of an unknown trial based on maximum likelihoods to various trained skill models. Using the parsed motion segments, correct classification of each occurrence of a surgeme showed that surgemes that were more commonly used were more indicative of skill. Taking a scoring of all the labels during a trial, surgeme-level models achieved 100% correct classification over our 57 datasets. The task-level HMM with unknown states correctly classified 94.7% of the trials. We further analyzed the task-level HMMs by comparing the "true" HMM built from unlabeled data with HMMs using labeled surgemes as states. Interestingly, we found that classification of trials using HMMs surgemes as states increased to 100%. This indicates that we would be able to input raw data into the evaluator without any prior labels and correctly classify skill almost as well.

There are several extensions of this work. More importantly, this method can correctly identify skill but does not answer how a novice model can move towards an expert model. The categorization of an intermediate surgeon is somewhat ambiguous since it is a class that is between an expert and novice. It is interesting to note that the analysis was able to detect what appears to be a slight learning curve in the intermediate data. Thus, this may be an area of future investigation to pinpoint where novices and experts differ during a trial and reinforce correct technique. We will need to better understand the surgeme representation to analyze which surgemes are the most important and most indicative of skill. In the near future, we intend to evaluate our methods with a larger, more variable database of surgical motions we have recently collected.

Acknowledgments. This material is based upon work supported by the National Science Foundation under Grant No. 0534359 and the Graduate Research Fellowship. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Moorthy, K., Munz, Y., Dosis, A., Hernandez, J., Martin, S., Bello, F., Rockall, T., Darzi, A.: Dexterity enhancement with robotic surgery. Surg. Endo. 18, 790–795 (2004)
2. Reznick, R., MacRae, H.: Teaching surgical skills-changes in the wind. New England Journal of Medicine 355(25), 2664–2669 (2006)
3. Xin, H., Zelek, J.S., Carnahan, H.: Laparoscopic surgery, perceptual limitations and force: A review. In: First Joint Eurohaptics Conf. and Symp. on Haptic Interfaces for Virtual Environment (2006)
4. Business, G.S.: Computer assisted surgical systems. A Global Strategic Business Report (2007)
5. Rabiner, L.: A tutorial on hidden markov models and selected applications in speech recognition. IEEE 77(2), 257–286 (1989)
6. Yang, J., Xu, Y., Chen, M.: Hidden markov model approach to skill learning and its application to telerobotics. Technical report, Robotics Institute, Carnegie Mellon University (1993)
7. Hannaford, B., Lee, P.: Hidden markov model analysis of force/torque information in telemanipulation. Int. Journal of Robotics Research 10(5), 528–539 (1991)
8. Hovland, G., Sikka, P., McCarragher, B.: Skill acquisition from human demonstration using a hidden markov model. Int. Conf. on Rob. Automat. 10(5), 528–539 (1996)
9. Yu, W., Dubey, R., Pernalete, N.: Robotic therapy for persons with disabilities using hidden markov model based skill learning. In: IEEE Int. Conf. on Rob. Automat., pp. 2074–2079 (2004)
10. Rosen, J., Brown, J.D., Chang, L., Hannaford, M.N.S.B.: Generalized approach for modeling minimally invasive surgery as a stochastic process using a discrete markov model. IEEE Trans. in. Bio. Eng. 53(3), 399–413 (2006)
11. Murphy, T.E., Vignes, C.M., Yuh, D.D., Okamura, A.: Automatic motion recognition and skill evaluation for dynamic tasks. In: Eurohaptics, pp. 363–373 (2003)
12. Megali, G., Sinigaglia, S., Tonet, O., Dario, P.: Modelling and evaluation of surgical performance using hidden markov models. IEEE Trans. on Bio. Eng. 53(10), 1911–1919 (2006)
13. Lin, H.C., Shafran, I., Yuh, D.D., Hager, G.D.: Towards automatic skill evaluation: Detection and segmentation of robot-assisted surgical motions. Comp. Aid. Surg. 11(5), 220–223 (2006)
14. Cao, C.G., MacKenzie, R.D., Payandeh, S.: Task and motion analyses in endoscopic surgery. ASME IMECE, 583–590 (1996)
15. Guthart, G.S., Salisbury, J.K.: The intuitivetm telesurgery system: Overview and application. In: IEEE Int. Conf. on Rob. Automat., pp. 618–621 (2000)
16. Hartigan, J.A., Wong, M.A.: A k-means clustering algorithm. Applied Statistics 8(1), 100–108 (1979)
17. Matlab: Statistics and pattern recognition toolkit

Development of the Ultra-Miniaturized Inertial Measurement Unit WB3 for Objective Skill Analysis and Assessment in Neurosurgery: Preliminary Results

Massimiliano Zecca[1,2,3,4], Salvatore Sessa[2], Zhuohua Lin[1], Takashi Suzuki[5], Tomoya Sasaki[6], Kazuko Itoh[4,7], Hiroshi Iseki[5], and Atsuo Takanishi[1,2,3,4,7,8]

[1] Global Robot Academia, Waseda University, Tokyo, Japan
zecca@aoni.waseda.jp
[2] Institute for Biomedical Engineering, ASMeW, Waseda University
[3] Humanoid Robotics Institute (HRI), Waseda University
[4] Italy-Japan Joint Laboratory on Humanoid and Personal Robotics "RoboCasa"
[5] FATS, ABMES, Tokyo Women's Medical University, Tokyo, Japan
[6] Graduate School of Science and Engineering, Waseda University
[7] Advanced Research Institute for Science and Engineering, Waseda University
[8] Department of Modern Mechanical Engineering, Waseda University

Abstract. In recent years there has been an ever increasing amount of research and development of technologies and methodologies aimed at improving the safety of advanced surgery. In this context, several training methods and metrics have been proposed, in particular for laparoscopy, both to improve the surgeon's abilities and also to assess her/his skills. For neurosurgery, however, the extremely small movements and sizes involved have prevented until now the development of similar methodologies and systems.

In this paper we present the development of the ultra-miniaturized Inertial Measurement Unit WB3 (at present the smallest, lightest, and best performing in the world) for practical application in neurosurgery as skill assessment tool. This paper presents the feasibility study for quantitative discrimination of movements of experienced surgeons and beginners in a simple pick and place scenario.

1 Introduction

In recent years, more and more technologies have entered the operating theatre, allowing surgeons to obtain better results and higher performance with great benefits for the patients by using cameras or microscopes and miniaturized tools. While these new technologies have many advantages for patients, they often require surgeons to undergo long and difficult training before achieving the necessary experience. One of the most important issues in this training is the objective evaluation and assessment of surgeon's dexterity and skill, and how they change and evolve over time.

Historically, this kind of evaluation has been subjectively performed by senior surgical staff in both training and operating environments. However it is fundamental to establish more efficient training exercises to enhance the dexterity of surgeons and to define objective metrics for assessing their experience and performance. While several metrics [1-3] and segmentation procedures [4-6] have been employed to characterize

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 443–450, 2009.

different phases of surgical movements in laparoscopy, the extremely small movements and sizes involved in neurosurgery (working field is usually 280-370 [mm^2] x h5-10[cm], target size 0.2-1.5[mm] [7]) have prevented until now the development of similar methodologies and systems.

Commercial systems such as StealthStation TREONplus by Medtronic or NDI's Aurora Electromagnetic Tracking System require either photoreflecive markers – by which position data are acquired at high frequency; however these markers are heavy and bulky, and they need to be always in sight with the digitizer, which is not always possible [8] – or EM tracking sensors – which are very small, but position data are acquired at relatively low freq, thus limiting the precision of the analysis. One possibility to overcome these limitations and to realize compact measurement systems is nowadays offered by Micro-Electro- Mechanical Systems (MEMS) technology. However, current prototypes such as WB2 [9] and commercial tools such as xSens MTx, InterSense InertiaCube3, and so on, are still too big for a practical application in neurosurgery.

Our aim, therefore, is to develop evaluation tools and to define a set of parameters that allow the characterization of the neurosurgeon's movements during surgery, to see how surgeons of different expertise performs, and to evaluate the improvement of the performance over training. In this paper we present the development of the ultra-miniaturized Inertial Measurement Unit WB3 and its application for the evaluation of surgeon's performance in a simple pick and place scenario.

2 Material and Methods

2.1 WB3 Inertial Measurement Unit IMU

Our group recently developed a new IMU which is very compact and lightweight (size 20 x 20 [mm] and weight 2.9 [g]) – at present the smallest, lightest, and best performing in the world. A picture of the new IMU is shown in Fig. 1. The IMU is composed by the following sensors: 3-axis accelerometer LIS3LV02DL; 2-axis gyroscope IDG300; 1-axis gyroscope LSIY300AL. The IMU's characteristics have been summarized in Table 1. This IMU actually also includes a 3-axis Magnetometer HMC5843; however the data of this sensor have not been analyzed in this paper, and therefore its description is omitted.

The LIS3LV02DL (STMicroelectronics) is a 3-axis accelerometer, whose small size (4.4 x 7.5 x 1 [mm]) and high performance characteristics are fully compatible with the strict requirements of neurosurgery. The resolution with a Full-scale = ±2 [g] and Bandwidth=160[Hz] is 2 [mg], with noise level of about 0.005 [m/sec^2] – less than one bit. The LISY300AL (STMicroelectronics) is a miniaturized 7.0 x 7.0 x 1.9 [mm] z-axis gyro sensor. Its full-scale is ±300 [deg/s] with a Bandwidth of 88 [Hz] and a sensibility of 3.3 [mV/deg/s]. In order to measure 3-axes angular velocities, we also used a bi-axial gyro IDG300 (InvenSense). The IDG300 size is 6.0×6.0×1.5 [mm], the measurement range is ± 500 [degree/s] and the sensitivity is 2.0 [mV/deg/s]. Unlike all other prototypes and commercial IMUs available today, this mixed configuration allows our IMU to obtain all the 3 axis of the gyros in one planar layer.

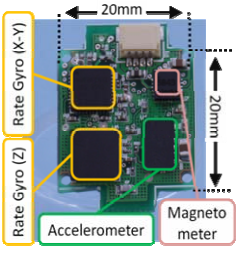


Fig. 1. The new Inertial Measurement Unit WB3

Table 1. Main characteristics of the inertial sensors in our IMU

	LIS3LV02DL	IDG300	LSIY300AL
Range	±2 [g]	±500[deg/s]	±300[deg/s]
Sensitivity	12±1bit	12±1bit	12±1bit
Bandwidth	40 Hz	140 Hz	88 Hz
Sample Rate	160 Hz	500 Hz	500 Hz
Linearity	±2%	<1%	±0.8%
Noise level	0.0051 m/s^2	0.35 [deg/s]	0.26 [deg/s]

Our IMU also contains a STMicroelectronics 32 bit microcontroller STM32 Cortex for embedded signal elaboration and data exchange. The communication with the module is performed using a CAN BUS at 1[Mb/s], directly connected with a PC.

2.2 Experimental Setup

During neurosurgery, one of the most commonly used instruments is the bipolar forceps (Fig. 2). The main characteristics of the system we used in our experiments are summarized in Table 2. A connector made by acrylonitrile butadiene styrene (ABS) polymer in rapid prototyping for housing our IMU is placed at the proximal end of the bipolar cutting tool (Fig. 2, right). Our IMU's extremely reduced weight and size allows it to be mounted on the bipolar forceps, and to be used during normal tasks without disturbing the surgeon's performance.

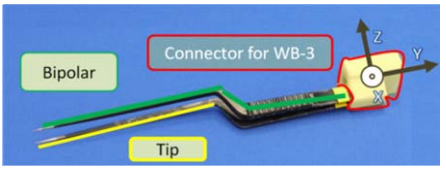


Fig. 2. The Bipolar cutting tool used in our experiments. X, Y, and Z axis are indicated

Table 2. Main characteristics of the Bipolar Cutting Tool

Bipolar Cutting Tool	
Total Length	194 [mm]
Tip Length	100 [mm]
Weight	34.0 [gr]

The Skill Evaluation System (SES) used for this preliminary experiment is shown in Fig. 3. SES is composed by 5 main parts:

1) The Testbed, made by ABS, simulates the most common operating space [7]. Current version has a size of 60 x 40 x h60 [mm].
2) The Support Base (SB) is an aluminium base of 100x100[mm] with a housing for the testbed made by ABS. SB's purpose is to hold the main unit stable.
3) The Surgical Field (SF, size 50x19 [mm]) simulates the aperture in the human skull. The surgeon accesses the Target Area and the Goal Area from here.
4) The Target Area (TA) is a replaceable soft surface made by Hitohada skin-like gel RTV-2K#1406 Hardness 0 (EXSEAL Corp., Tokyo, Japan), on which the targets are placed for the experiment. Three different types of TAs, each with 5 small targets randomly placed on it, were prepared to simulate the typical objects that are handled

during neurosurgery (Fig. 3(b)): BIG: 3.2 x 1.6 x 0.4 [mm]; MEDIUM: 2.0 x 1.2 x 0.3 [mm]; SMALL: 1.0 x 0.5 x 0.2 [mm]. Five target areas for each type were prepared in advance to simplify the experimental procedure.
5) The targets picked up from the TA are put in the Goal Area (GA, size 10 x 30 [mm]). To facilitate the release of the target, a putty-like adhesive is placed at the bottom of this area.

The experimental setup is shown in Fig. 3(c). The microscope is a MITAKA MRI (Mitaka Kohki Co., Ltd, Tokyo, Japan).

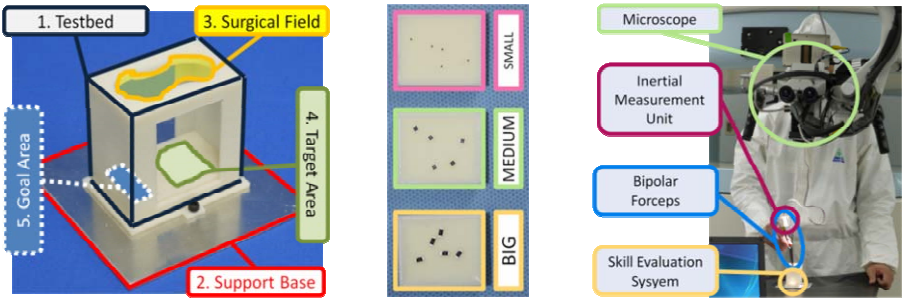


Fig. 3. (a) Picture of the Skill Evaluation System (SES); (b): The 3 types of target area; (c): Overview of the experimental setup

2.3 Experimental Protocol

Thirteen non-medical subjects (average age 27.53 years, age range 22-39, all male, all right handed), and 1 professional neurosurgeon (male, age 40, right handed) kindly agreed to participate to the experiments after providing informed consent. Among the non-medical subjects, only 1 had some experience with neurosurgical tools, and 1 had some experience with laparoscopy; all the other subjects were totally novice. The experiments consist in picking all the targets in the target area, one by one, with the bipolar forceps, and releasing them in the goal area. The target area is replaced at the end of each exercise. In total there are 15 target areas (3 sizes x 5 repetitions), and they are replaced by following the order BIG → MEDIUM → SMALL, for 5 times.

Acceleration data were sampled at f_{acc} =160 [Hz]; gyro data at f_{gyro} = 500 [Hz]. Data were acquired on the pc for real-time display, and saved for storage and offline analysis. All the data were saved as CSV (Comma Separated Value), and then loaded in Matlab™ (The MathWorks, Inc.) for further processing and analysis. For each repetition the data were then automatically trimmed to remove dead-time at the beginning and at the end of the trial due to the manual start/stop, as follows: n_{start} is defined as the first sample when $|a_y| > 9.1\,m/s^2$; n_{end} is defined as the first sample from the end when $|a_y| > 9.1\,m/s^2$. (Y is the long axis of the bipolar cutting tool as defined in Fig. 2). Only samples between n_{start} and n_{end} were then analyzed.

Acceleration components and angular speed components were then filtered and smoothed by using a 10th order bandpass IIR Butterworth filter with cutoff frequencies f_{c1} = 0.05 Hz, f_{c2} = 8hz (accelerometers), f_{c1} = 0.05 Hz, f_{c2} = 5 Hz (rate gyros), to remove bias and to remove physiological tremor [10].

In this paper we present the calculation and analysis of the following variables: Execution Time T_{task}; Acceleration module $|a|$; and Angular Speed module $|\omega|$. The fft for $|a|$ was calculated with $FFT_{size} = 8192$ samples and frequency resolution $f_{res(acc)}=f_{acc}/ FFT_{size} = 160/8192 = 0.0195$ [Hz]. The fft for $|\omega|$ was calculated with frequency resolution $f_{res(gyro)}=f_{gyro}/ FFT_{size} = 0.061$ [Hz]. The Power Spectral Density PSD was estimated in both cases with the following formula: $(\text{fft} \cdot \text{conj}(\text{fft}))/\text{fft}_{\text{size}}$. The frequency range chosen for the evaluation was 0.2 – 8 [Hz] for $\text{fft}(|a|)$, and 0.2 – 2 [Hz] for $\text{fft}(|\omega|)$, to take into account only the voluntary movements. Several other parameters were also calculated, but are not presented here due to space limitation.

3 Experimental Evaluation

The following sections present the details about the experimental evaluation. In the following figures, *(norm.)* indicates that the data have been normalized to the average corresponding data of the surgeon (Subject #14) for an easier visual comparison of the scales. The normalizing values are always indicated. Surgeon's data are always displayed with dark gray bars, while novice's data are displayed in light grey. Moreover, the y-scales of the normalized graphs are optimized in function of the surgeon's value, which means that some bar exceeds the y-scale limit (i. e. Fig. 6).

3.1 Analysis of the Execution Time

The surgeon – as expected – proved to be always faster (i.e. lower T_{task}) and showed high constancy in the execution time (i.e. lower $\text{var}(T_{task})$) of the different tasks than all the novices (Fig. 4). In addition, it can be noticed that both BIG (Fig. 5, left) and MEDIUM (not shown) targets s howed a fast learning effect for the novices, with the execution time stabilizing after the 3rd trial; however, these learning effects could not be seen for the SMALL target (Fig. 5, right). It's also worth noticing that this learning effect is also present with the surgeon for his first trials (not shown).

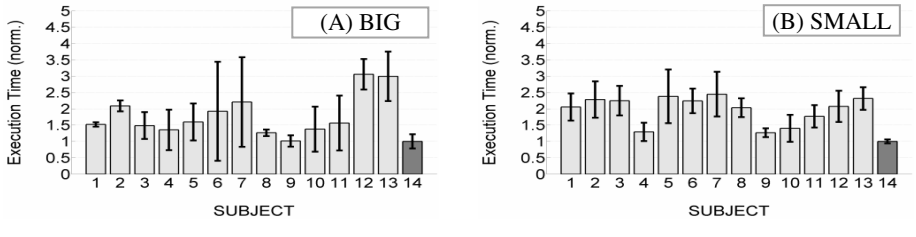


Fig. 4. Execution time for the BIG (left) and for the SMALL (right) targets, averaged on the 5 trials for each subject. Normalization T_{task} values are 11.75 and 16.17 [s], respectively

3.2 Analysis of the Acceleration

As can be seen in Fig. 6, the mean Power Spectral Density $\overline{PSD_{|a|}}$ clearly identifies the experienced neurosurgeon (for the surgeon it is 0.62 and 1.47 $[(m/s^2)^2)]$ for the BIG and SMALL task, respectively; while for the other subjects it's usually twice as high or

more). Frequency distribution of $PSD_{|a|}$ is also quite different. It can also be seen that subjects #2 and #9 performed very similarly to the neurosurgeon.

Another parameter which shows some differences between the surgeon and the novices is the Cumulative Distribution Function CDF of the acceleration module $|a|$, calculated as $CDF(|a|) = P(X \leq |a|)$ and evaluated for $X = 95\%$ ($CDF_{95\%}$). For the BIG targe t (Fig. 7 (left)) the neurosurgeon outperforms all the novices; the data about the MEDIUM target shows similar trends. Things are different for the SMALL target, when several novices have similar or lower $CDF_{95\%}(|a|)$ than the neurosurgeon. These however are due to a much higher T_{task} for the novices.

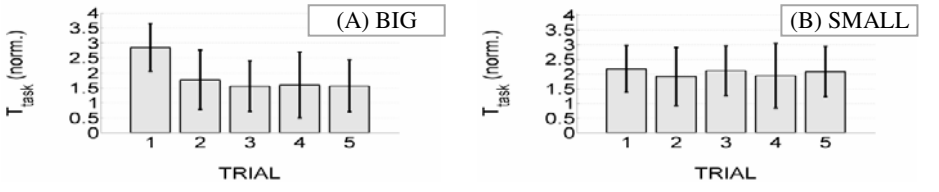


Fig. 5. Execution time for the BIG (left) and for the SMALL (right) targets averaged on all non-medical subjects for each trial. Normalization T_{task} are 11.75 and 16.17 [s], respectively.

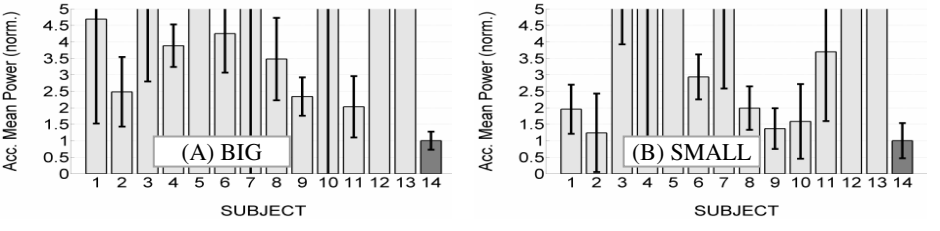


Fig. 6. PSD for the (A) BIG and for the (B) SMALL targets, averaged on the 5 trials for each subject. Normalization values are 0.62 and 1.47 [$(m/s^2)^2$)].

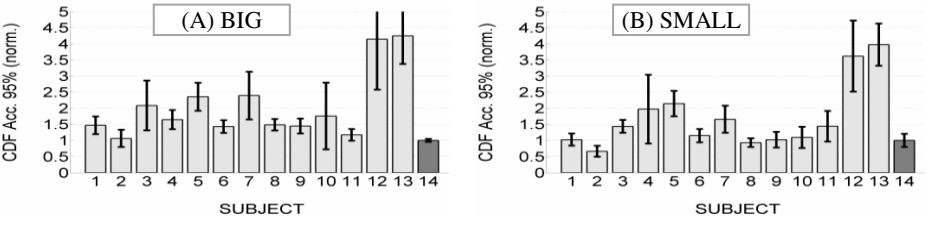


Fig. 7. Cumulative Distribution Function (CDF) of the acceleration $|a|$; for the (A) BIG and (B) SMALL targets. Normalization values are 1.10 and 1.38 [m/s^2], respectively.

3.3 Analysis of the Rate Gyro

Among the different parameters, the Cumulative Distribution Function (CDF) of the angular speed $CDF(|\omega|) = P(X \leq |\omega|)$ evaluated for $X = 95\%$ ($CDF_{95\%}$) shows some

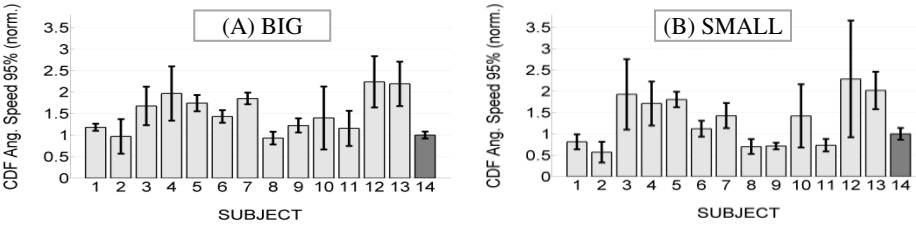


Fig. 8. CDF of the angular speed. Normalization values are 16.6 and 19.3 [deg/s].

difference between the surgeon and the n ovices (Fig. 8). As with $CDF_{95\%}(|a|)$,, the surgeon's $CDF_{95\%}(|\omega|)$ is usually lower than the other subjects for the BIGGER and MEDIUM targets, showing therefore high smoothness in movement; it's intermediate for the SMALL targets, where some subject reached a lower $CDF_{95\%}(|\omega|)$ at the expenses of a higher T_{task}. $CDF_{95\%}(|\omega|)$ also shows a very limited variance for the surgeon, thus signifying high regularity in the exercises.

4 Discussion and Conclusions

With the diffusion of more and more advanced tools and technologies in the operating room, it is fundamental to establish more efficient training exercises and to define objective metrics to objectively evaluate the dexterity of neurosurgeons. The extremely small movements and sizes involved in neurosurgery, however, have prevented until now the development of such methodologies and systems. In this paper we presented the development of an ultra-miniaturized Inertial Measurement Unit WB3 (Sec. 2.A). The extreme lightweight and its extreme high performance make WB3 suitable for applications in neurosurgery. Moreover, it directly acquires accelerations and rotations at high frequency, thus providing smooth and virtually noise-free data. Commercial systems, instead, acquire target's position and then obtain rotations and accelerations by successive derivations which introduce more noise.

As a preliminary test we applied WB3 to a bipolar forceps (Sec. 2), the most commonly used instrument in neurosurgery, and we used this system to evaluate the movements of neurosurgeons in a simple pick and place scenario. The preliminary results (Sec. 3) proved that several parameters extracted from the IMU's data (and in particular the mean Power Spectral Density $\overline{PSD_{|a|}}$ and the Cumulative Distribution Function $CDF_{95\%}$ of both acceleration $|a|$ and angular speed $|\omega|$) allow a clear distinction between a professional neurosurgeon and a group of novices; moreover, these data also could show which non-medical subject performs similarly to the surgeon, and how, thus validating the approach proposed in this pilot study.

In the future, the data of more professional subjects, as well as the data of both more non-medical and more medical novice subjects are needed to define a more precise evaluation system. Overall, this work shows that a substantial set of parameters is necessary to investigate and analyze the performance of surgeons. Currently, work is still in progress, and our future commitment in this field is to continue to analyze the performance of neurosurgeons in more complex exercises and procedures.

Acknowledgments. This research has been conducted at the Waseda University Center for Advanced Biomedical Sciences/TWIns. Partial support was also provided by: ASMeW Priority Research C Grant #11; JSPS Grant-in-aid for Scientific Research #19700389 and #20700386; JSPS Postdoctoral Fellowship Program for Foreigner Researcher FY2008; G-COE Global Robot Academia; Advancement of University Education Project of Chinese government; Waseda Grant for Special Research Projects #266740. The authors would like to express their thanks to the Italian Ministry of Foreign Affairs, General Directorate for Cultural Promotion and Cooperation, for its support to RoboCasa. The authors would also like to express their gratitude to STMicroelectronics, Dyden Corp, Japan ROBOTECH LTD, SolidWorks Corp, Okino Industries LTD, for their support to the research.

References

1. Cotin, S., Stylopoulos, N., Ottensmeyer, M.P., Neumann, P., Rattner, D.W., Dawson, S.L.: Metrics for laparoscopic skills trainers: The weakest link! In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 35–43. Springer, Heidelberg (2002)
2. Verner, L., Oleynikov, D., Holtmann, S., Haider, H., Zhukov, L.: Measurements of the level of surgical expertise using flight path analysis from da VinciTM robotic surgical system. Medicine Meets Virtual Reality 11, 373–378 (2003)
3. Stylopoulos, N., Vosburgh, K.G.: Assessing technical skill in surgery and endoscopy: a set of metrics and an algorithm to assess skills in surgical and endoscopic procedures. Surg. Innov. 14, 113–121 (2007)
4. Risucci, D., Cohen, J.A., Garbus, J.E., Goldstein, M., Cohen, M.G.: The effects of practice and instruction on speed and accuracy during resident acquisition of simulated laparoscopic skills. Curr. Surg. 58, 230–235 (2001)
5. Payandeh, S., Lomax, A., Dill, J., Mackenzie, C., Cao, C.G.L.: On Defining Metrics for Assessing Laparoscopic Surgical Skills in a Training Envirnoment. Studies in Health Technology and Informatics - Medicine Meets Virtual Reality, vol. 85, pp. 334–340. IOS Press, Amsterdam (2002)
6. Megali, G., Sinigaglia, S., Tonet, O., Dario, P.: Modelling and evaluation of surgical performance using hidden Markov models. IEEE transactions on bio-medical engineering 53, 1911–1919 (2006)
7. Gonzalez, L.F., Crawford, N.R., Horgan, M.A., Deshmukh, P., Zabramski, J.M., Spetzler, R.F.: Working area and angle of attack in three cranial base approaches: pterional, orbitozygomatic, and maxillary extension of the orbitozygomatic approach. Neurosurgery 50, 550–555; discussion 555–557 (2002)
8. Woerdeman, P.A., Willems, P.W., Noordmans, H.J., van der Sprenkel, J.W.: The analysis of intraoperative neurosurgical instrument movement using a navigation log-file. Int. J. Med. Robot. 2, 139–145 (2006)
9. Zecca, M., Saito, M., Endo, N., Mizoguchi, Y., Itoh, K., Takanobu, H., Takanishi, A.: Development of the Waseda Bioinstrumentation System WB-2 - the new Inertial Measurement Unit. In: IEEE International Conference on Robotics and Biomimetics (Robio 2007), Sanya, China, pp. 139–144 (2007)
10. Ibanez, D.O., Baquerin, F.P., Choi, D.Y., Riviere, C.N.: Performance Envelope and Physiological Tremor in Microsurgery. In: Proceedings of the IEEE 32nd Annual Bioengineering Conference, pp. 121–122 (2006)

Novel Endoscope System with Plasma Flushing for Off-Pump Cardiac Surgery

Ken Masamune[1], Tetsuya Horiuchi[1], Masahiro Mizutani[1], Hiromasa Yamashita[1], Hiroyuki Tsukihara[2], Noboru Motomura[2], Shinichi Takamoto[2], Hongen Liao[3], and Takeyoshi Dohi[1]

[1] Graduate School of Information Science and Technology, The University of Tokyo
[2] Graduate School of Medicine, The University of Tokyo
[3] Graduate School of Engineering, The University of Tokyo
masa@i.u-tokyo.ac.jp

Abstract. The purpose of this study is to develop a new endoscope for performing simple surgical tasks inside a cardiac atrium/chamber filled with blood, i.e., for performing "off-pump" cardiac surgeries. In general, it is very difficult to observe the inner wall of the vessels containing circulating blood because the light from the endoscope is scattered by the red blood cells. "Plasma flushing" performed using the separator system is developed to observe the inner side of the heart filled with blood and to remove blood cells from the front of the endoscope tip. The system was used in in vitro quantitative measurement of the device performance and in vivo experiments on a swine. In these experiments, we successfully obtained high-resolution images of the interior of the heart during off-pump surgery.

1 Introduction

Currently, the cardiopulmonary bypass system is being widely used for performing cardiotomy under extracorporeal circulation even though the operation task is quite simple—the operation involves simply clipping valve leaflets, cutting or shortening a chord, closing the septal hole, etc. However, there are risks associated with the use of an artificial heart-lung machine during the operation. For example, complications such as cerebral infarct and pulmonary infarct may arise after operation. Thus, in the case of simple surgery inside the heart, off-pump surgery is ideal for increasing the quality of life of the patients.

Some research groups and companies have been trying to perform surgical tasks by using off-pumps, e.g., Linte CA[3] developed the augmented reality environment for off-pump mitral valve implantation, and L. Ming[4] developed an MRI-compatible robot for assisting off-pump artificial aortic valve replacement. Tomasz [1] and Evalve Inc.[2] reported the development of the surgical catheter device "Mitraclip" that can be used to perform off-pump edge-to-edge (E2E) repair for mitral regurgitation; in this procedure, the obstruction of transmitral flow is reduced by attaching clips to the leaflets of the mitral valve through catheter manipulation. However, it is very difficult to perform quick, efficient operations by adopting these procedures or using these devices because of the lack of image information during surgery. For each

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 451–458, 2009.

task, surgeons have to use low-resolution ultrasound images, 2D X-ray images obtained by X-ray exposure, or delayed MR images in order to observe the procedure during the surgery. The best observation method is to obtain an endoscopic view of the interior of the heart. Endovascular endoscopes are widely used not only for diagnosis but also for the surgical treatment of vessel stenosis, vessel sclerosis, etc. However, it is very difficult to observe the inner side of the vessels containing circulating blood because the light from the endoscope is scattered by the red blood cells and clear images are not obtained if saline flushing of these vessels is not performed.

To overcome the above problems, we developed a new endoscope system to acquire a clear endoscopic view that would help perform simple surgical tasks inside a cardiac atrium/chamber filled with blood. We utilized a "plasma flushing" system to remove blood cells from the front of the endoscope tip, as described in section 2. A new tip design is adopted for more efficient observation. In section 3, experiments and their results are described. In contrast, when using conventional endovascular endoscopes, an innovative method must be developed for observation of the inner surface of a wider heart chamber, given the strong turbulence caused by heart beats.

2 Design and Development of New Endoscope

To observe the inner chamber of the heart filled with blood, we developed a "plasma flushing" system to remove blood cells from the front of the endoscope tip. Figure 1 shows an overview of our prototype system. The system comprises a whole endoscope (The diameter and length of the image guide fiber are 3.8 mm and 200 mm, respectively.) with a small channel (d = 2.0 mm); this channel comprises a coaxial

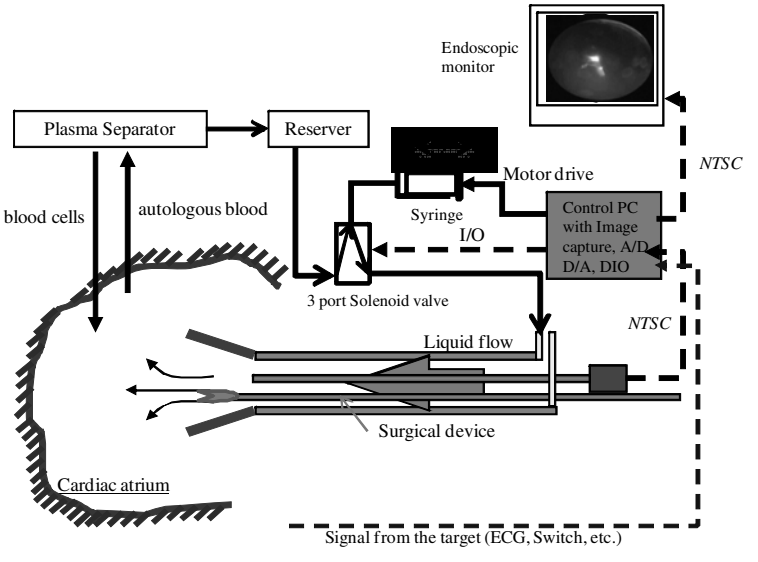


Fig. 1. Overview of the plasma flushing system: a fiber endoscope, surgical device, and the liquid flow are integrated inside the endoscope device (d = 10 mm)

flush channel and a small channel for forceps; the system also comprises a three-port solenoid valve, a syringe with a pumping actuator, a control PC with an add-in board for capturing images, A/D, D/A, and DIO extension boards to control the timing of the flushing, and an observation monitor.

The required specifications for the endoscope are as follows:

1. The external diameter should be less than 10 mm.
2. The cardiovascular function should not be disturbed.
3. Points up to 10 mm from the endoscope tip should be observable.
4. The endoscope must be biocompatible and sterilizable.

Figure 2 shows the prototype design of the device that satisfies the above requirements. Inside the outer pipe, a fiber-optic endoscope (d = 0.7 mm; AS-003, FiberTech Inc.) and biopsy forceps (d = 1.7 mm) are placed; further, the ends of the pipe are sealed by silicone rubber. Inside of the outer pipe, d = 0.7 mm fiber-optic endoscope (AS-003, FiberTech Inc.) and d = 1.7 mm biopsy forceps are set and the end of pipe are sealed by silicone rubber. A conventional endoscope can only be used to observe blood flow in a beating heart because blood contains corpuscles that reflect and scatter light. To observe intracardiac surgery, we used a clear liquid flushed from the endoscope tip created a transparent region to observe the object. We proposed the use of plasma from autologous blood as the flushing liquid, because it is transparent, viscous, fully biocompatible, and if the total amount of liquid can be controlled, it will not cause stress to the heart during the operation.

We attached the new tool called the "separator," which looks like an umbrella, to separate liquid and blood at the device tip. The separator is made from a thin plastic film and six frames that are made of SUS304. The frames are arranged in a radial pattern around the center of the endoscope. This separator can be folded by pushing the outer sheath, and it recovers its original shape by the action of a self-generated spring force.

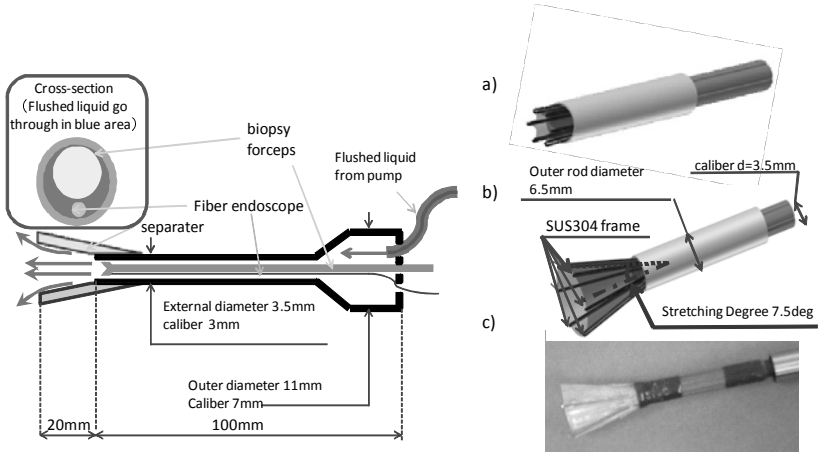


Fig. 2. Detailed design of the endoscopic device with separator tip: a) Status before device insertion b) Opening of the separator c) Developed separator tip

3 Experiments and Results

3.1 Evaluation Experiment for Observable Depth

We performed a phantom experiment to evaluate the relationship between the angle of the separator of the tip and the observable depth. The experimental procedure is as follows:

1. Prepare a tank filled with blood and place a plastic phantom inside the tank.
2. Fix the endoscope device on an XYZ stage to measure the distance between the device tip and the phantom.
3. Measure the maximum depth upto which the phantom can be clearly observed.
4. Vary the following parameters: the separator open-angle, the distance, and the flushing speed.

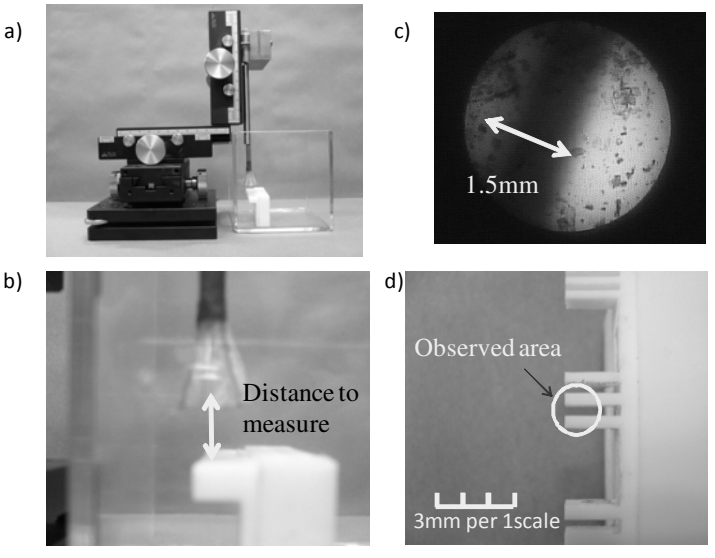


Fig. 3. Experimental setup: a) Endoscopic device attached on an XYZ stage along with the phantom; b) The distance to measure; c) Endoscope image obtained during flushing performed inside the blood pool; d) Observed area of the phantom

Table 1. Relationship between separator angle and measured observation depth

Separator angle[deg]	Flush speed [ml/s]	Observable depth [mm]
13	5.4	2.5
7.5	2.4	3.8
0	4.8	3.5

In this experiment, we used glycerin as a flushing liquid instead of blood plasma.

The results are listed in Table 1. The listed flushing speeds are the maximum speeds required to observe the target phantom. The observable depth was the best when the open angle of the separator was set to 7.5°.

3.2 Evaluation Experiment for Observation Area

The observation area (visual field) was evaluated using a phantom along with a device that was equipped with a separator and one that was not equipped with a separator. In this experiment, the separator with an open angle of 7.5° was used since it gave the best performance in the previous experiment. We used another phantom, shown in Figure 4, to measure the circular observation area with a diameter of 5 mm. We evaluated the maximum depth by varying the distance between the endoscope tip and the phantom while the entire circle was in view.

The results are listed in Table 2. It was impossible to observe the phantom at any distance when the endoscope was used without a separator.

Table 2. Dependence of the maximum observable depth on the separator

Separator	Observation depth [mm]	Flushing speed [ml/s]
Present	3	4
Absent	Not Visible	At any speed

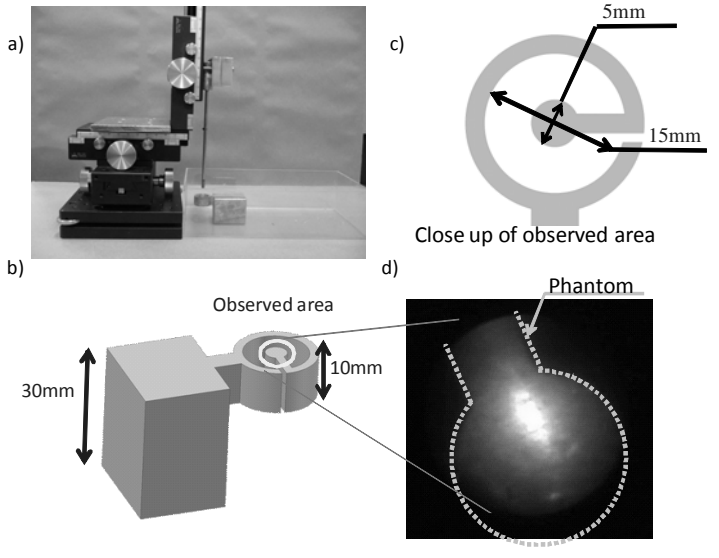


Fig. 4. Experimental setup: a) Endoscopic device attached on an XYZ stage and the phantom; b) Phantom for measuring the observable area; c) Dimensions of the phantom; d) Observed phantom during the flushing performed in the blood pool

3.3 In vivo Experiments and Results

In vivo experiments were performed on a swine whose heart size is similar to the size of a human heart. In the experiment, the device was inserted into the right atrium through a trocar with a diameter of 4.0 mm; the trocar was placed in an off-pump.

We chose the tip of tricuspid valve and chorda tendinea in the right atrium for observation. Since the anatomy of the swine heart is different from that of the human heart, the trocar-insertion approach from the left atrium is very severe. Thus, we chose the right atrium for the in vivo experiment. The observation was monitored by real-time ultrasound imaging, which was required to recognize the position of the device tip before performing the flushing in the right atrium. This is because a clear view was available in an extremely narrow space and the surgeon could not find the correct anatomical structures without the position and orientation information.

We also used glycerin as a flushing liquid instead of blood plasma because of the restriction on the experiment time. The glycerin liquid was warmed to maintain the swine's body temperature at a constant level so that no shocks were felt by the heart. The flushing speed during the experiment was 3–5 ml/s.

The screenshots from the video obtained during the off-pump observation are shown in Figure 5. The image on the left is the normal image that corresponds to the case without flushing - we can only observe blood flow. The image at the center shows the chordae tendineae of the tricuspid valve, and the papillary muscle was also observed well for a few seconds. The image on the right clearly shows the leaflet of the tricuspid valve in the right atrium.

In these experiments, we successfully obtained very clear camera images of the interior of the beating heart during off-pump surgery.

4 Discussion

In in vitro experiments, we used real plasma, and the results were not very different from the in vivo results. However, some problems related to crystallization of the plasma were encountered, and we need to consider the temperature control.

The in vitro experiment was performed in the stationary condition, and the experimental results showed a trade-off between the observable depth and observable area.

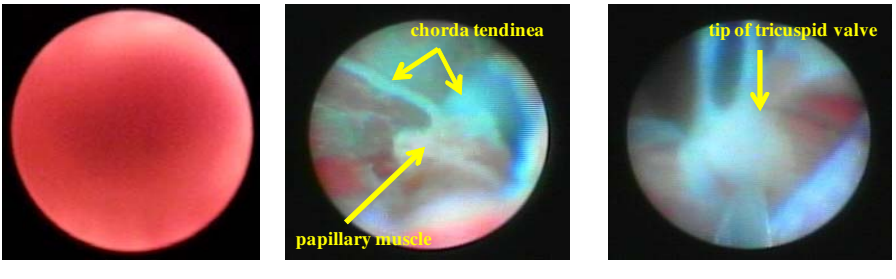


Fig. 5. In-vivo experiment: (Left) Observation without flushing—we observe nothing because of the presence of blood; (Center) Observation when flushing was performed—Chorda tendinea and papillary muscle were observed; (Right) The tip of the tricuspid valve was observed

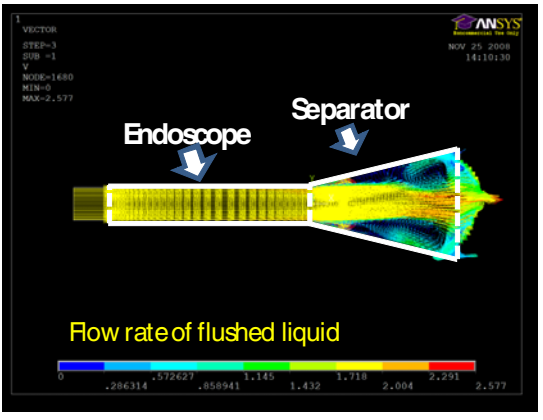


Fig. 6. 2D flow simulation of the device performed using ANSYS: the white lines represent the boundary, and the bar to the bottom represents the flow rate of the flushing liquid

We believe that this trade-off arises because the observable volume is constant for any flow rate while the observable depth varies with flow rate.

However, an actual clinical case involves the dynamic state, and further analysis is required for detailed design of the device. In a preliminary study, we attempted to perform 2D flow simulations using ANSYS CFD, as shown in Figure 6. In this simulation, the turbulent separated flow inside the separator flows along the separator wall, which is a boundary. Further, it is expected that in this simulation, it is possible to obtain a wider transparent volume for a high-velocity flow at the device tip. However, it is very difficult for us to perform an analysis by simulating the 3D incompressible unsteady flow. As the next step, we should carry out both parameter acquisitions by performing in vivo experiments and further high-cost simulations.

In the in vivo experiment, we developed a PC system that can be used to control the pumping timing through a manual switch or ECG signals; the system can also be used to store the endoscope images in real-time and periodically display images (similar to a stroboscopic view) on the monitor in order to maintain a clear view that is not distorted by the presence of blood cells. However, because of the time delay in the liquid flushing, it was impossible to control the timing of the flushing in the in vivo experiment. The flush pumping mechanism requires a few improvements before it is used in the next task.

In all cases, we successfully obtained clear endoscope images of the interior of the heart, even though each observation lasted for a short time. The next task is to combine surgical tools to perform the operation. N Suzuki has already developed the same kind of endoscope with a small robot; the endoscope can be used for intravascular operation[5].

5 Conclusion

In this study, we developed a new endoscope system to acquire clear endoscopic images of the interiors of blood vessels or the heart. These images are useful when

performing simple procedures during cardiac surgery. By using this device, we can obtain high-resolution images to observe the inner structure of the heart and its pumping activity. Thus, we can perform better diagnosis and administer more effective therapy by applying robotic technologies inside the heart.

We are currently developing a small manipulator that we will combine with the observation device for performing advanced minimal invasive surgery in the near future.

This research is supported by MEXT Grant-in-Aid for Scientific Research (S) (#17100008).

References

1. Timek, T.A., Nielsen, S.L., Lai, D.T., Tibaya, F., Liang, D., Daughters, G.T., Beineke, P., Hastie, T., Ingels Jr., N.B., Miller, D. C.: Mitral Annular Size Predicts Alfieri Stitch Tension in Mitral Edge-to-Edge Repair. J. Heart Valve Dis. 13(2), 165–173 (2004)
2. `http://www.evalveinc.com/index.html` (March 2009)
3. Linte, C.A., Wiles, A.D., Moore, N.H.J., Wedlake, C., Gerard, G., Jones, D., Bainbridge, D., Peters, T.M.: An augmented reality environment for image-guidance of off-pump mitral valve implantation. In: Proc. SPIE International Symposium on Medical Imaging 2007, San Diego, CA, February 17-22, 2007, pp. 6509–6522 (2007)
4. Li, M., Mazilu, D., Horvath, K.A.: Robotic System for Transapical Aortic Valve Replacement with MRI Guidance. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 476–484. Springer, Heidelberg (2008)
5. Suzuki, N., Hattori, A., Suzuki, S., Otake, Y.: Development of a surgical robot system for endovascular surgery with augmented reality function. In: Medicine Meets Virtual Reality, vol. 15, pp. 460–463. IOS Press, Amsterdam (2007)
6. Masamune, K., Mizutani, M., Tsukihara, H., Takamoto, S., Matsumiya, K., Dohi, T.: A new blood-permeable endoscope system with plasma flushing for off-pump cardiac surgery. International Journal of Computer Assisted Radiology and Surgery 3(suppl. 1), S245–S246 (2008)

Endoscopic Orientation Correction

Kurt Höller[1,⋆], Jochen Penne[1,⋆⋆], Armin Schneider[2], Jasper Jahn[3], Javier Guttiérrez Boronat[3], Thomas Wittenberg[1,3], Hubertus Feußner[2], and Joachim Hornegger[1]

[1] Chair of Pattern Recognition (LME) and Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander University Erlangen-Nuremberg
hoeller@informatik.uni-erlangen.de
[2] Workgroup for Minimal Invasive Surgery (MITI), Klinikum rechts der Isar, Technische Universität München, Germany
[3] Fraunhofer-Institute for Integrated Circuits IIS, Erlangen, Germany

Abstract. An open problem in endoscopic surgery (especially with flexible endoscopes) is the absence of a stable horizon in endoscopic images. With our "Endorientation" approach image rotation correction, even in non-rigid endoscopic surgery (particularly NOTES), can be realized with a tiny MEMS tri-axial inertial sensor placed on the tip of an endoscope. It measures the impact of gravity on each of the three orthogonal accelerometer axes. After an initial calibration and filtering of these three values the rotation angle is estimated directly. Achievable repetition rate is above the usual endoscopic video frame rate of 30Hz; accuracy is about one degree. The image rotation is performed in real-time by digitally rotating the analog endoscopic video signal. Improvements and benefits have been evaluated in animal studies: Coordination of different instruments and estimation of tissue behavior regarding gravity related deformation and movement was rated to be much more intuitive with a stable horizon on endoscopic images.

1 Introduction

In the past years, *N*atural *O*rifice *T*ranslumenal *E*ndoscopic *S*urgery (NOTES) [1] has become one of the greatest new challenges within surgical procedures and has the strong potential to eventually succeed minimal invasive surgery (MIS). Currently, MIS interventions are mainly carried out by surgeons using *rigid* laparoscopes inserted in the abdomen from the outside, while gastroenterologists apply *flexible* video-endoscopes for the detection and removal of lesions in the

⋆ I am thankful to G. Hager, P. Kazanzides, R. Kumar, D. Mirota and H. Girgis for their helpful suggestions in the preparation of this manuscript during my research stay at the Engineering Research Center for Computer-Integrated Surgical Systems and Technology, The Johns Hopkins University, Baltimore.

⋆⋆ The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 459–466, 2009.

gastro digestive tract (esophagus, stomach, colon, etc.). As the currently practiced *NOTES* and *hybrid* interventions require flexible endoscopes to access the abdominal cavity as well as the surgical instruments and skills to perform the actual intervention, both disciplines and technologies are needed. Gastroenterologists have been trained and accustomed to navigate through the lumen of the colon, stomach or esophagus by pushing, pulling and rotating the flexible video-endoscope (fig. 1), regardless of orientation, rotation and pitch of the endoscope tip inside the patient and the image orientation displayed on the monitor. Surgeons, on the other hand, are used to a fixed relation between the tip of the endoscope and the inside of the patient, as neither one of them is changing their position during the intervention. However, mismatches in the spatial orientation between the visual display space and the physical workspace lead to a reduced surgical performance [2,3].

Hence, in order to assist surgeons interpreting and reading images from flexible video-endoscopy, an automated image rectification or re-orientation according to a pre-defined main axis is desirable [4]. The problem of the rotated image is even more important in hybrid NOTES procedures, where an additional micro instrument is inserted through the abdominal wall for exposition and tasks during extremely complex interventions.

In the past, there have been suggested different approaches for motion tracking [5] and image rectification [6]. Several approaches use parameters achieved from registration of intra-operative obtained 3-D data with pre-operative CT or MRI volumes. Such intra-operative 3-D data can be obtained from image-driven approaches like monocular shape-from-shading [7] and structure-from-motion [8,9], stereocular triangulation [10], active illumination with structured light [11] or application of an additional time-of-flight/photonic-mixing-device camera [12]. But even if intra-operative 3-D data can be obtained and reconstructed in real-time, e.g. via time-of-flight cameras needing no data post-processing and having frame rates higher than 30Hz, real-time computation of registration parameters is still a challenge [13] especially since colon or stomach provide less applicable feature points.

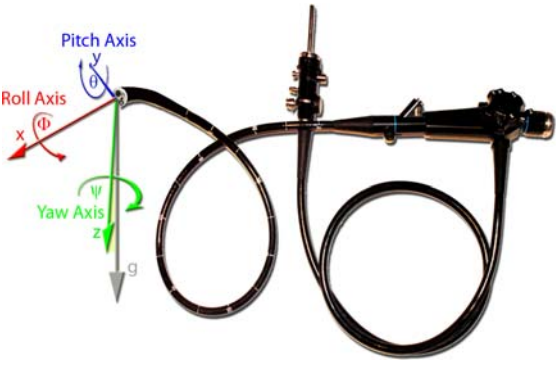


Fig. 1. Roll, pitch and yaw description for endoscopic orientation

A broad overview of possible tracking technologies has been given in [5]. These also include the idea of electro-magnetic tracking, which can be applied to an endoscope. This requires not only an additional sensor in the endoscope's tip but also an external magnetic field. This can easily be disturbed by metallic instruments and leads to several further restrictions [14]. A by far simpler approach to measure the needed orientation angle will be presented in this work and consists of integrating a Micro Electro-Mechanical System (MEMS) based inertial sensor device in the endoscope's tip to measure influencing forces in three orthogonal directions (fig. 1). If the endoscope is not moving, only the acceleration of gravity has an effect on the three axes.

2 Method

2.1 Technical Approach

To describe the orientation of the endoscope relating to the direction of gravity, an Cartesian "endoscopic board navigation system" with axes $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{z}$ (according to the DIN 9300 aeronautical standard [15]) is used as body reference frame [16]. The tip points in x-direction which is the boresight, the image bottom is in z-direction and the y-axis is orthogonal to both in horizontal image direction to the right. Rotations about these axes are called roll Φ (about x), pitch Θ (about y) and yaw Ψ (about z). Image rotation has only to be performed about the optical axis x which is orthogonal to the image plane. Gravity $\boldsymbol{g}$ is considered as an external independent vector. Since there is no explicit angle information, only the impact of gravity on each axis can be used to correct the image orientation. Equation (1) expresses, how rotation parameters Φ, Θ and Ψ of the IMU (Inertial Measurement Unit) have to be chosen to get back to a corrected spatial orientation with $\boldsymbol{z}$ parallel to $\boldsymbol{g}$:

$$\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\Phi) & \sin(\Phi) \\ 0 & -\sin(\Phi) & \cos(\Phi) \end{pmatrix} \cdot \begin{pmatrix} \cos(\Theta) & 0 & -\sin(\Theta) \\ 0 & 1 & 0 \\ \sin(\Theta) & 0 & \cos(\Theta) \end{pmatrix} \cdot \cdot \begin{pmatrix} \cos(\Psi) & \sin(\Psi) & 0 \\ -\sin(\Psi) & \cos(\Psi) & 0 \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 0 \\ g \end{pmatrix} = \begin{pmatrix} -\sin(\Theta)g \\ \sin(\Phi)\cos(\Theta)g \\ \cos(\Phi)\cos(\Theta)g \end{pmatrix} \tag{1}$$

with $F_{x,y,z}$: measured acceleration

Using the two-argument function arctan2 to handle the arctan ambiguity within a range of $\pm\pi$ one finally can compute roll Φ for $F_x \neq \pm g$ and pitch Θ for all values:

$$\Phi = arctan2(F_y, F_z) \tag{2}$$

$$\Theta = \arcsin\left(\frac{-F_x}{g}\right) \tag{3}$$

As g determines just 2 degrees of freedom with this approach yaw Ψ cannot be computed. If $F_x = \pm\, g\ (\ \rightarrow\ \Theta = \pm\pi\ \rightarrow\ F_y = F_z = 0\ )$ roll Φ is not determinable either. To avoid movement influence, correction is only applied if superposed acceleration additional to gravity g is below boundary value ΔF_{absmax}:

$$|\sqrt{F_x^2 + F_y^2 + F_z^2} - g| < \Delta F_{absmax} \tag{4}$$

First, a preceded 3×3 calibration matrix, which incorporates misalignment and scaling errors [17,18], has to be retrieved by initial measurements. Moreover a peak elimination is the result of down sampling the measuring frequency, which is considerably higher than the image frame rate (up to 400Hz vs. 30Hz). This is realized by summing up separately all n sensor values F_{x_i}, F_{y_i} and F_{z_i} within an image frame with $i = 1, ..., n$ and weighting them with a weighting factor w_i with maximal weight w_0:

$$w_i = \frac{1}{\frac{1}{w_0} + |\sqrt{F_{x_i}^2 + F_{y_i}^2 + F_{z_i}^2} - g|} \tag{5}$$

Afterwards the sum has to be normalized by the sum of all weighting factors w_i:

$$\begin{pmatrix} F_x \\ F_y \\ F_z \end{pmatrix} = \sum_{i=1}^{n} \left(\begin{pmatrix} F_{x_i} \\ F_{y_i} \\ F_{z_i} \end{pmatrix} \cdot w_i \right) \cdot \sum_{i=1}^{n} (w_i)^{-1} \tag{6}$$

To avoid bouncing or jittering images as a result of the angle correction, additional filtering is necessary. Hence, prior to angle calculation, each axis is filtered with a Hann filter to smooth angle changes and with a minimum variation threshold ΔF_{axmin} to suppress dithering. As long as superposed acceleration calculated in equation (4) remains below boundary value ΔF_{absmax}, roll Φ and pitch Θ can be calculated using equations (2) and (3). Otherwise they are frozen untill ΔF_{absmax} is reached again. If these boundaries are chosen correctly, the results will be continuous and reliable since nearly all superposed movements within usual surgery will not discontinue or distort angle estimation. Both original and rotated image are

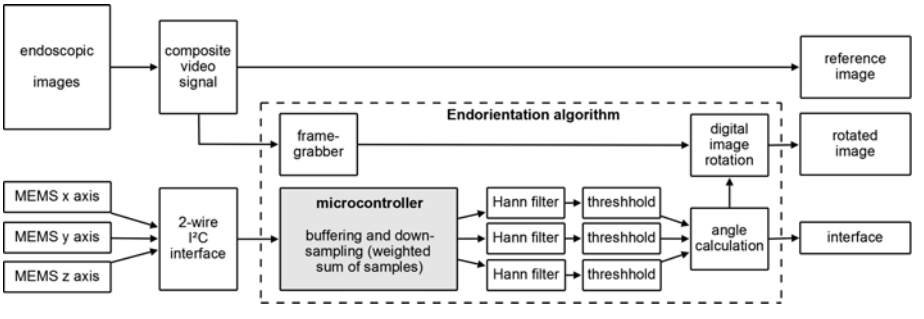


Fig. 2. Block diagram of rotation correction with the "Endorientation" algorithm

displayed for security reasons. For potential use with other devices the calculated angle is also transmitted to an external communication interface (fig. 2).

2.2 Image Rotation

The measurement data is transferred as a digital signal via a two-wire I^2C interface along the flexible endoscope tube. The endoscopic video signal is digitalized via an external USB video capture device with an adequate resolution to provide the usual quality to the operator. By this design the "Endorientation" algorithm is divided into two parts. One part running on a small 8-Bit microcontroller and one parting running as an application on a workstation. Everytime the capture device acquires a new frame the software running on the workstation requests the actual acceleration values from the software on the microcontroller. The three acceleration values are used to calculate the rotation angle according to the equations above. The rotation of the frame is performed via the OpenGL library GLUT [19]. The advantage of this concept is the easy handling of time-critical tasks in the software. We can use the sensor sample rate of 400Hz doing some filtering without getting into trouble with the scheduler granularity of the workstation OS. The information of the endoscope tip attitude is available within less than 30ms. Our "Endorientation" approach can be performed in real-time on any off-the-shelf Linux or Windows XP/Vista workstation.

2.3 Clinical Evaluation

In a porcine animal study, the navigation complexity of a hybrid endoscopic instrument during a NOTES peritoneoscopy with the well established transsigmoidal access [20] was compared with and without Endorientation. The endoscopic inertial measurement unit was fixed on the tip of a flexible endoscope (fig. 3). Additionally a pulsed DC magnetic tracking sensor was fixed on the hybrid instrument holder for recording the position of the surgeon's hands. To

Fig. 3. Prototyping with external sensor on the endoscope's tip

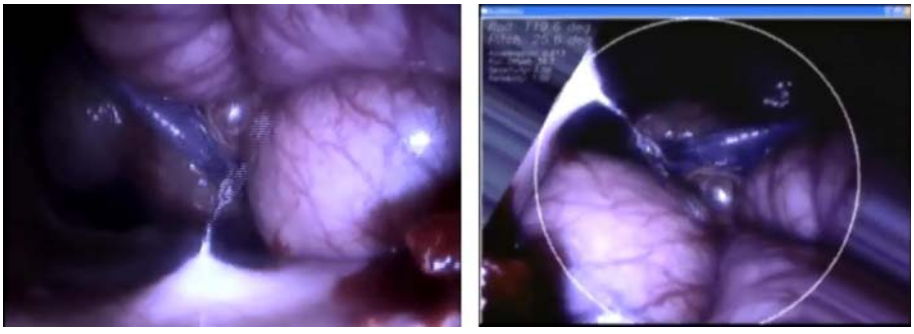


Fig. 4. Original (l) and rotated (r) image with needle incision and fluid injection

evaluate the benefit of automated MEMS based image rectification, four different needle markers were inserted through the abdominal wall to the upper left and right and the lower left and right quadrants. Under standardized conditions these four needle markers had to be grasped with a trans-abdominal introduced endoscopic needle holder. Displaying alternately originally rotated and automatically rectified images path and duration were recorded and analyzed.

3 Results

3.1 Technical Accuracy

With the employed sensor there is a uniform quantization of 8 bit for a range of ±2.3g for each axis. This implies a quantization accuracy of 0.018g per step or 110 steps for the focused range of ±g. This is high enough to achieve a durable accuracy even to a degree within relatively calm movements. This is possible as roll angle Φ is calculated out of inverse trigonometric values of two orthogonal axes. Single extraordinary disturbed MEMS values are suppressed by low weighting factors w_i. Acceleration occurs only in the short moment of changing movement's velocity or direction. For the special case of acceleration with the same order of magnitude as gravity, ΔF_{absmax} can be chosen small enough to suppress calculation and freeze the angle for this short period of time. By choosing a longer delay line for the smoothing Hann filter and a higher minimum variation threshold ΔF_{axmin}, correction may be delayed by fractions of a second but will be stable even during fast movements.

3.2 Clinical Evaluation

In the performed experiments, it could clearly be shown that grasping a needle marker with an automatically rectified image is much more easier and therefore faster than with the originally rotated endoscopic view (fig. 4). In comparison to the procedure without rectification the movements are significantly more accurate with by factor 2 shorter paths and nearly half the duration. The details of

verified clinical benefits are described in [21]. Obviously the two parameters *duration* and *path length* are strongly correlated and can be regarded as a significant measure for the complexity of surgical procedures. Since both are decreased with the application of image rectification, the complexity of the complete procedure can be reduced.

4 Discussion

As described in the previous section, an automatic rectification (or re-orientation) of the acquired endoscopic images in real-time assists the viewer in interpreting the rotated pictures obtained from a flexible videoscope. This is especially important for physicians, who are used to naturally rectified endoscopic images related to a patient-oriented Cartesian coordinate system within their surgical site. In contrast, gastroenterologists have learned by combination of long experience, anatomical knowledge and spatial sense how to use and interpret an endoscope-centered (tube-like) coordinate system during their exploration of lumenal structures, even if the displayed images are rotating. Our described experiments included surgeons originally unrelated to flexible endoscopes. For future research, we will also include gastroenterologists, who are experienced reading and interpreting rotated and non-rectified image sequences. Possibly, in the future of NOTES, dual monitor systems will be needed to support both specialists during the intervention.

The hardware costs for of-the-shelf communication converter, capture device, MEMS sensor and circuit board are below $250. More reliable hardware will increase this amount by some factors, Linux/XP/Vista workstation and an additional Display have to be added. Integrating the sensor board in a flexible endoscope is surely possible but probably more expensive as well. However, there are several two channel endoscopes available. One of their working channels could be used for the MEMS sensor. In conclusion, we have shown that it is simple and affordable to additionally provide rotated images with fixed horizon for better orientation. The main complexity while using the second working channel could be to fix the sensor in the lumen, to prevent rotation in the working channel and to get the possibility to change the sensor with an instrument.

References

1. Rattner, D., Kalloo, A.: ASGE/SAGES working group on Natural Orifice Translumenal Endoscopic Surgery: White Paper. Surg. Endosc. 20, 329–333 (2006)
2. Holden, J., Flach, J., Donchin, Y.: Perceptual-motor coordination in an endoscopic surgery simulation. Surg. Endosc. 13, 127–132 (1999)
3. Cao, C.G., Milgram, P.: Disorientation in minimal access surgery: A case study. In: Proceedings of the IEA 2000/HFES 2000 Congress, vol. 4, pp. 169–172 (2000)
4. Koppel, D., Wang, Y.-F., Lee, H.: Automated image rectification in video-endoscopy. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 1412–1414. Springer, Heidelberg (2001)
5. Welch, G., Foxlin, E.: Motion tracking: No silver bullet, but a respectable arsenal. IEEE Comput. Graph. Appl. 22(6), 24–38 (2002)

6. Koppel, D., Wang, Y.F., Lee, H.: Robust and real-time image stabilization and rectification. In: Procs. 7th IEEE Workshop on Application of Computer Vision (WACV/MOTION 2005), vol. 1, pp. 350–355. IEEE Computer Society Press, Los Alamitos (2005)
7. Yeung, S.Y., Tsui, H.T., Yim, A.: Global shape from shading for an endoscope image. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 328–332. Springer, Heidelberg (1999)
8. Deguchi, K., Sasano, T., Arai, H., Yoshikawa, H.: 3D shape reconstruction from endoscope image sequences by the factorization method. IEICE Transactions on Information and Systems 79(9), 1329–1336 (1996)
9. Thormählen, T., Broszio, H., Meier, P.N.: Three-dimensional endoscopy. In: Falk Symposium No. 124, Medical Imaging in Gastroenterology and Hepatology, Hannover, September 2001, vol. 124 (2002)
10. Stoyanov, D., Darzi, A., Yang, G.Z.: A practical approach towards accurate dense 3D depth recovery for robotic laparoscopic surgery. Computer Aided Surgery 4(10) (June 2005)
11. Albitar, C., Graebling, P., Doignon, C.: Fast 3D vision with robust structured light coding. In: SPIE Medical Imaging 2009: Visualization and Image-Guided Procedures, Orlando, USA (February 2009)
12. Penne, J., Höller, K., Krüger, S., Feußner, H.: NOTES 3D: Endoscopes learn to see 3D; basic algorithms for a novel endoscope. In: Araújo, A.H., Vitriá, H.J. (eds.) Proceedings of VISAPP 2007, pp. 134–139 (2007)
13. Mirota, D., Taylor, R.H., Ishii, M., Hager, G.D.: Direct endoscopic video registration for sinus surgery. In: Medical Imaging 2009: Visualization, Image-guided Procedures and Modeling. Proceedings of the SPIE, February 2009, vol. 7261 (2009)
14. Hummel, J., Figl, M., Kollmann, C., Bergmann, H., Birkfellner, W.: Evaluation of a miniature electromagnetic position tracker. Med. Phys. 29(10), 2205–2212 (2002)
15. DIN 9300-1: Aerospace; concepts, quantities and symbols for flight dynamics; aircraft motion relative to the air; ISO 1151-1:1988 modified. Deutsches Institut Fuer Normung e.V (German National Standard) (October 1990)
16. Titterton, D., Weston, J.: Strapdown Inertial Navigation Technology, 2nd edn. John Wiley and Sons, Chichester (2001)
17. Dorobantu, R.: Simulation des Verhaltens einer low-cost Strapdown IMU unter Laborbedingungen. Schriftenreihe des IAPG (1999)
18. Höller, K.: Characterisation and modeling of an inertial sensor for navigation of autonomous systems. Diploma thesis, Friedrich-Alexander University Erlangen-Nuremberg (October 2005)
19. Kilgard, M.J.: The OpenGL Utility Toolkit (GLUT) Programming Interface API Version 3. Silicon Graphics, Inc. (1996)
20. Wilhelm, D., Meining, A., von Delius, S., et al.: An innovative, safe and sterile sigmoid access (ISSA) for NOTES. Endoscopy 39, 401–406 (2007)
21. Höller, K., Schneider, A., Jahn, J., Guttierrez, J., Wittenberg, T., Hornegger, J., Feussner, H.: Clinical evaluation of Endorientation: Gravity related rectification for endoscopic images. In: Proceedings of the ISPA 2009 (in press, 2009)

Time-of-Flight 3-D Endoscopy*

Jochen Penne[1], Kurt Höller[1], Michael Stürmer[1], Thomas Schrauder[2], Armin Schneider[3], Rainer Engelbrecht[2], Hubertus Feußner[3], Bernhard Schmauss[2], and Joachim Hornegger[1]

[1] Chair of Pattern Recognition and Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander-University Erlangen-Nuremberg, Germany
Jochen.Penne@informatik.uni-erlangen.de
[2] Chair of Microwave Engineering and High Frequency Technology, Friedrich-Alexander-University Erlangen-Nuremberg, Germany
[3] Workgroup for Minimally Invasive Therapy and Intervention, Klinikum rechts der Isar, Technical University Munich

Abstract. This paper describes the first accomplishment of the Time-of-Flight (ToF) measurement principle via endoscope optics. The applicability of the approach is verified by in-vitro experiments. Off-the-shelf ToF camera sensors enable the per-pixel, on-chip, real-time, marker-less acquisition of distance information. The transfer of the emerging ToF measurement technique to endoscope optics is the basis for a new generation of ToF rigid or flexible 3-D endoscopes. No modification of the endoscope optic itself is necessary as only an enhancement of illumination unit and image sensors is necessary. The major contribution of this paper is threefold: First, the accomplishment of the ToF measurement principle via endoscope optics; second, the development and validation of a complete calibration and post-processing routine; third, accomplishment of extensive in-vitro experiments. Currently, a depth measurement precision of 0.89 mm at 20 fps with 3072 3-D points is achieved.

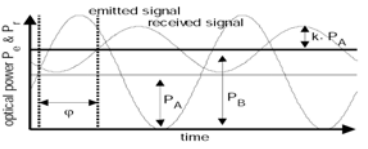
1 Introduction

Minimally invasive surgery (MIS) has become a promising option if not standard procedure for a great number of surgical interventions due to the minimized collateral surgical trauma and quicker recovery. Minimally invasive intervention techniques had, have and will have a significant impact on both patients and surgeons.

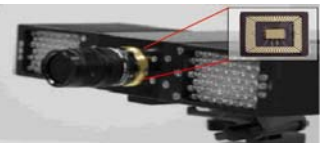
Today diagnosis and intervention are based on the inspection of an image sequence acquired by a monocular video camera in MIS. This image sequence provides a flat visualization of the operation area. Consequently, depth perception and navigation of surgical instruments is very difficult. The recognition and assessment of pathological structures and tissues as well as the quantification

* The authors gratefully acknowledge funding of the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 467–474, 2009.



(a) ToF measurement principle



(b) ToF camera system

Fig. 1. Fig. 1(a) illustrates the ToF measurement principle. The phase-delay φ between the sinusoidally modulated illumination light power P_e and received light power P_r is caused by the light propagation delay over the distance. φ is measured in each pixel of a ToF sensor. The incoming amplitude $k \cdot P_A$ ($0 \leq k \leq 1$ depending on signal attenuation) is also measured by ToF sensors and provides a grey-scale image. (P_B: optical power of background light sources.) Fig. 1(b) shows a commercially available ToF camera system (PMD[vision]3k-S from PMDTec GmbH) and the 7.4 mm×6.4 mm ToF sensor located inside the housing. To the left and right the infrared active illumination units can be observed. Replacing the original lense by an endoscope optic and replacing the original illumination unit by a customized one were the basic hardware modification steps to build the ToF endoscope.

of their three-dimensional dimensions is only possible due to the expertise and experience of the inspecting person and is consequently subjective in its nature. By three-dimensional surface models of the operation area such tasks can be accomplished more easily, more objective and in a reproducible manner. It is a widely investigated and verified fact that 3-D information significantly improves operative safety and precision during intervention [1] as well as during surgical training [2], for example in laparoscopic or gastrointestinal endoscopy [3].

No widely applicable and easy-to-build system is available which enables real-time, *3-D measuring* (in contrast to *3D visualization*) at a constant pixel resolution in the operation area during an endoscopic intervention. The system proposed in this article provides this measuring capability with sub-millimeter depth precision.

2 State of the Art

Intra-operatively acquiring depth information is a problem which has gained significant interest especially in the field of MIS.

Image-driven monocular or stereoscopic approaches for recovering depth data of the operation area like [4,5] have been proposed. In general, these techniques unfortunately provide no real-time capability as they require the processing of an image sequence; Furthermore, they provide no guaranteed density of the computed 3D point cloud as the number of computed 3D points relies on the amount of detected and tracked features. Approaches have been proposed which weaken assumptions like a static field-of-view or taking care of missing data problems [6,7]. Such image-driven approaches have been successfully applied for surgical robot-safety management or autonomous positioning of surgical instruments.

Instead of applying vision techniques the utilization of measuring techniques for directly acquiring the depth information from the operation area have been

investigated. Utilizing a miniaturized digital holography system 512×512 3D points can be obtained at 5 fps [8]. The authors successfully address non-medical application fields but do not comment on the applicability under sterile conditions. This is an important issue as the proposed measurement probe is mounted at the tip of the endoscope and thus has to be sterilized for intra-operative usage. In [9] and [10] an approach based on the detection of a laser beam line which is actively controlled by an optical galvano scanner is described. Approx. 4000 3D points can be obtained at 5-6 fps which was verified to be enough information for a robotic navigation system. The approach requires the insertion of two monocular endoscope optics: One for the projection of the laser beam and one for observing the projected laser beam. The effort of inserting an additional endoscope optic to obtain the 3D geometry may not be suitable for practical clinical use.

Our approach aims at utilizing a single endoscope optic, which is commonly available during endoscopic interventions, to derive 3D information of the operation area.

A measurement principle, which can derive 3D information if only one endoscope optic is available, is the ToF measurement principle. ToF sensors have been developed without a focus on MIS or endoscopes and are available since one decade [11]. ToF sensors consist of a pixel matrix and an external illumination unit, which actively illuminates the scene with an incoherent near infrared light. This light is intensity-modulated with a modulation frequency f_{mod}, which is usually in the range of $\leq$30 MHz for commercially available ToF sensors. Each pixel is synchronized with the illumination unit and measures the phase-delay φ due to the propagation delay between the emitted and reflected light. This approach is illustrated in Fig. 1(a). The phase-delay is related to the propagation time of the signal t_d by $t_d = \varphi/(2\pi \cdot f_{mod})$, and the traveled distance d of the signal can be computed by $d = c \cdot t_d$, where c is the speed of light in the transmission medium ($c \approx 299.710 km/s$ in air). Additionally, each pixel provides an amplitude information proportional to the intensity of the incoming infrared light. The amplitude information is mainly depending on the distance and the reflecting material. This information provides a gray-value image (amplitude data image) in addition to the distance map.

3D Cartesian coordinates can be computed from the measured distances. Standard calibration routines [12] can be applied to compute the necessary intrinsic camera parameters if they are not known in advance.

3 Method and Evaluation

3.1 Illumination Unit of the ToF Endoscope

For the ToF principle a light source with fast intensity modulation is required. As this modulation frequency f_{mod} is in the range of 10 MHz to 100 MHz a mechanical modulation of the light beam by a rotating chopper wheel can not be used. Thermal light sources like tungsten incandescent lamps used typically for endoscope illumination can not be modulated that fast by the electrical

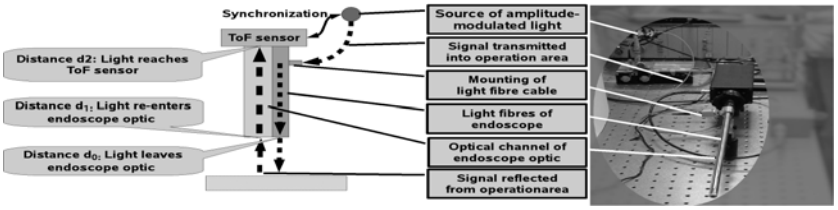


Fig. 2. Scheme of the signal transmission chain and hardware setup of a ToF endoscope. Note that $d_0 \leq d_1 < d_2$. The large dashed line depicts the transmission of the illuminating light from the light source to the operation area. The small dashed line depicts the transmission of the illuminating light from the operation area to the ToF sensor.

current, either. Therefore, light emitting diodes (LEDs) are typically applied for the illumination unit of ToF cameras as can be seen in Fig. 1(b). LEDs can be modulated by their electrical current up to 100 MHz. However, to generate a sufficient light intensity, many LEDs are required in parallel. This makes such an illumination unit impractical for coupling to the illumination fiber guide of an endoscope. For this work, a single fiber-coupled high-power laser diode has been adopted to the endoscope. With an output power of max. 2 W emitted from a single 200 μm diameter optical fiber, this laser diode can easily be coupled to the endoscope and provides sufficient light power to overcome the transmission losses of the endoscope illumination and image guides for a good signal-to-noise ratio of the ToF camera. The high-frequency characteristics of the laser diode have been studied thoroughly to design the required high-speed driver electronics. By a single RF MOSFET transistor the modulation of the laser diode up to frequencies of 50 MHz with potential to reach 100 MHz for future ToF cameras with an improved distance resolution was enabled. The modulation is synchronized with the ToF camera for accurate phase measurements. Thus, a powerful and versatile illumination light source for adopting standard 3D ToF cameras to endoscopes was realized.

3.2 Data Processing

As Fig. 2 depicts, the distance value (if not explicitly stated otherwise all distances are specified in mm) computed by a pixel of the ToF sensor is not initially the distance from the endoscope tip to the operation area, but rather biased by a constant error $d_0 + (d_2 - d_1)$.

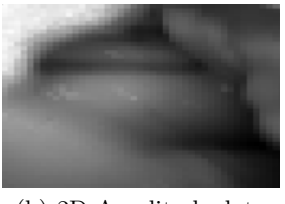
The distance $d_0 + (d_2 - d_1)$ is different for each pixel due to slight differences in the transmission way through the fiber optics and optical channel of the endoscope optic. To remove this error an opaque object of good reflectivity (for example a sheet of white paper) is held directly before the endoscope optic: The distance measured in each pixel corresponds to $d_0 + (d_2 - d_1) = d_2$ as $d_1 - d_0 = 0$. The acquired distances are additionally post-processed with a bilateral filter using a spatial sigma of 10.0 and a range sigma of 50.0 [13] to get

a smooth 2D distance-correction mask. Subtracting the computed offset mask from the acquired distance map and dividing the obtained values by two yields the distances of the observed points from the endoscope tip. Such distances will be further on referred to as *offset-corrected*. To reduce outliers a 2D bilateral filter (with spatial sigma 1.0 and range sigma 5.0) is finally applied to the offset-corrected distance map. Pixels whose distance measurement is severely corrupted either by acquiring very few of the infrared light or by acquiring too much of it (which leads to saturation effects in a pixel and consequently to an invalid distance measurement) can be automatically identified by applying an upper and lower threshold to the amplitude data. As the amplitude information is an absolute measurement of the acquired amplitude of the observed infrared signal a global threshold can be chosen.

3.3 Evaluation of Measurement Precision and In-vitro Experiments

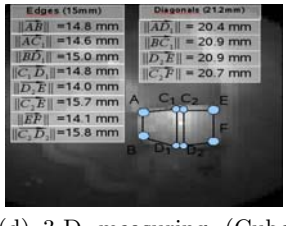
A ToF endoscope utilizing the ToF sensor of the ToF camera PMD[vision]3k-S, which operated at a modulation frequency of 30 MHz, and a zero degree endoscope optic with 10 mm diameter was used for the tests and experiments. The ToF sensor had a lateral resolution of 64×48 pixels and operated at upto

(a) In-vitro experimental setup

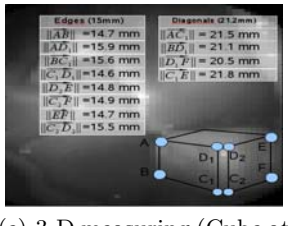


(b) 2D Amplitude data

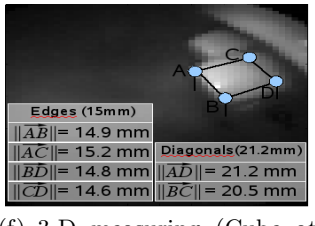
(c) Depth map



(d) 3-D measuring (Cube at 4.6 cm; scene range: 1.4 cm-13 cm)

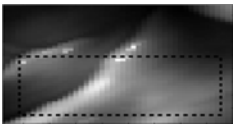


(e) 3-D measuring (Cube at 3 cm; scene range: 1.2 cm-8 cm)

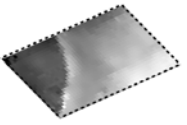


(f) 3-D measuring (Cube at 4 cm; scene range: 2.5 cm-10 cm)

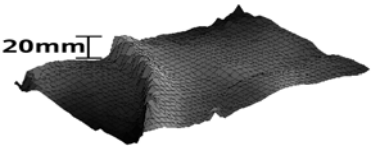
Fig. 3. Fig. 3(a) shows the setup used for the experiments. Fig. 3(b)-3(c) show the data provided by the ToF endoscope: A 2D amplitude data image (Fig. 3(b)) and a distance map (Fig. 3(c): the brighter a pixel the closer is the point) of a tunnel-like anatomical structure at the entrance of the stomach. Fig. 3(d)-3(f): Endoscopic 3-D measurement of edge and surface diagonal length of a 15×15×15 mm cube from different viewing positions in the porcine stomach. The depicted gray value image encodes the amplitude data (not the distances) provided by the ToF endoscope.



(a) Amplitude image of gastric mucosa



(b) Distance map of ROI



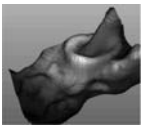
(c) 3D Visualization of ROI

(d) Object dimension

(e) 2D Color image

(f) ToF amplitude image

(g) ToF distance map



(h) 3D Visualization

Fig. 4. Various example data sets. **Porcine stomach experiments:** Fig. 4(a): Amplitude data with Region-of-Interest(ROI); Fig. 4(b): Distance map of ROI; Fig. 4(c): 3D visualization. **2D endoscopic color data vs. ToF endoscopic 3D data:** A ring-shaped object (Fig. 4(d)) was inserted into a pepper (Fig. 4(e): 2D endoscopic color image). ToF endoscope data: Amplitude data (Fig. 4(f)), depth map (Fig. 4(g)), 3D Visualization (Fig. 4(h)).

25 fps. If not explicitly stated otherwise, the data (acquired using the ToF sensor) which is referred to in the following has been acquired by a calibrated endoscope optic and the distances have been offset-corrected. For a working distance of 3 cm the measurement precision was computed. The measurement expression is expressed for each pixel by the standard deviation of the acquired distances when observing a static scene. An average precision of 0.89 mm and a median precision of 0.71 mm was computed from 100 acquired distance maps. A lower amplitude threshold of 50 and an upper threshold of 700 were chosen to detect pixels with a severely corrupted distance measurement (see section 3.2). 5% of the pixels were neglected. An empty porcine stomach was manually insufflated with air. The ToF endoscope was inserted via the remaining parts of the esophagus. Before inserting the endoscope optic two plastic cubes each of size $15 \times 15 \times 15$ mm were inserted into the stomach. This served the purpose to be able to observe three-dimensionally objects of known size and shape. During the experiments one of the cubes was observed from different viewing positions. The four 3-D boundary points of each visible surface square were manually selected. The length of the edges (ground truth: 15 mm) of the square as well as the length of its diagonals (ground truth: 21.2 mm) were computed. The 3-D point coordinates of the selected points were utilized. Thus, the computed lengths are 3D Euclidean distances between the selected 3-D points. The results along with illustrative examples of the amplitude and distance data provided by a ToF endoscope are depicted in Fig. 3. A 3D surface reconstruction of the stomach mucosa is given in Fig. 4 along with a comparison of 2D endoscopic color data and ToF endoscopic 3D data.

4 Discussion and Conclusion

The hardware costs for turning an available endoscope optic into a ToF endoscope utilizing only commercially available components are less than 5000 Euro ($7000). The ToF endoscope operates at 20 fps including data acquisition and the complete processing chain (dual-core 2.4 GHz PC, 2GB RAM). The mean measurement precision of 0.89 mm is sufficient to provide valuable intra-operative information. The immediately available distance map of the operation area enables a direct visual depth observation. The capability to three-dimensionally measure anatomical structures (tumors, etc.) in the operation area was verified by using an artificially inserted cube of known dimensions (see Fig. 3(d)-3(f)).

ToF endoscopes introduce outstanding perspectives to minimally invasive surgery. Approaches addressing collision detection, robot-guided surgery, intra-operative navigation support and 3-D visualization of the operation area will benefit from the proposed novel 3-D endoscope. Additionally, the capability to measure distances and dimensions of user-selected anatomical structures three-dimensionally in real-time in the operation area has the capability to provide valuable diagnostic information to the surgeon. Innovative new endoscopic intervention techniques like NOTES (Natural Orifice Transluminal Endoscopic Surgery) may be accomplished more successfully by utilizing ToF endoscopes. Considering the ToF measurement principle and endoscope optics there are no restrictions: Any rigid or non-rigid endoscope optic can be utilized as a ToF endoscope in the proposed manner. Only chip-on-tip endoscopes can not be used in such a way.

5 Outlook

Currently, the utilized infrared illumination unit does not meet the safety requirements for being utilized outside a controlled laboratory. An illumination operating in the visible spectral range would provide two advantages: a better quantum efficiency of ToF sensors and the eye-safety requirements are somewhat less restrictive in the range of visible light.

Next generation ToF sensors will provide lateral resolutions of 204×204 pixels. The utilization of such sensors for a ToF endoscopes will improve the observability of smaller anatomical structures. By mounting a beam-divider and an additional standard CCD image sensor, 2D color information and distance information can be acquired with one endoscope optic. The presented experiments have been accomplished with a rigid endoscope optic. A quantitative evaluation of the capabilities of a flexible ToF endoscope is subject to current research.

References

1. Wengert, C., Bossard, L., Häberling, A., Baur, C., Székely, G., Cattin, P.C.: Endoscopic Navigation for Minimally Invasive Suturing. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 620–627. Springer, Heidelberg (2007)

2. Votanopoulos, K., Brunicardi, F., Thornby, J., Bellows, C.: Impact of three-dimensional vision in laparoscopic training. World Journal Of Surgery 32(1), 110–118 (January 2008)
3. Yoshida, T., Inoue, H., Hara, E., Umezawa, A., Ohtsuka, K., Endo, S., Tamegai, Y., Kashida, H., Tanaka, J., Kudo, S.: Newly developed 3D endoscopic system: preliminary experience. Endoscopy 35(2), 181–184 (2003)
4. Burschka, D., Li, M., Taylor, R., Hager, G.D.: Scale-Invariant Registratiou of Monocular Endoscopic Images to CT-Scans for Sinus Surgery. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 413–421. Springer, Heidelberg (2004)
5. Mountney, P., Stoyanov, D., Davison, A.J., Yang, G.-Z.: Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 347–354. Springer, Heidelberg (2006)
6. Hu, M., Penney, G., Edwards, P., Figl, M., Hawkes, D.: 3D Reconstruction of Internal Organ Surfaces for Minimal Invasive Surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 68–77. Springer, Heidelberg (2007)
7. Lo, B., Scarzanella, M.V., Stoyanov, D., Yang, G.Z.: Belief Propagation for Depth Cue Fusion in Minimally Invasive Surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 104–112. Springer, Heidelberg (2008)
8. Kolenovic, E., Osten, W., Klattenhoff, R., Lai, S., von Kopylow, C., Jüptner, W.: Miniaturized Digital Holography Sensor for Distal Three-Dimensional Endoscopy. Appl. Opt. 42(25), 5167–5172 (2003)
9. Hayashibe, M., Suzuki, N., Nakamura, Y.: Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management. Medical Image Analysis 10(4), 509–519 (2006); Special Issue on Functional Imaging and Modelling of the Heart (FIMH 2005)
10. Hayashibe, M., Suzuki, N., Hattori, A., Nakamura, Y.: Intraoperative Fast 3D Shape Recovery of Abdominal Organs in Laparoscopy. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2489, pp. 356–363. Springer, Heidelberg (2002)
11. Xu, Z., Schwarte, R., Heinol, H., Buxbaum, B., Ringbeck, T.: Smart Pixel – Photometric Mixer Device (PMD) / New System Concept of a 3D-Imaging-on-a-Chip. In: 5^{th} International Conference on Mechatronics and Machine Vision in Practice, Nanjing, pp. 259–264 (1998)
12. Zhang, Z.: A Flexible New Technique For Camera Calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)
13. Tomasi, C., Manduchi, R.: Bilateral Filtering for Gray and Color Images. In: Sixth International Conference on Computer Vision (ICCV 1998), pp. 839–846 (1998)

In Vivo OCT Coronary Imaging Augmented with Stent Reendothelialization Score

Florian Dubuisson[1], Claude Kauffmann[2], Pascal Motreff[1,3], and Laurent Sarry[1]

[1] ERIM, Faculty of Medicine, University of Auvergne, 63001 Clermont-Ferrand, France
{florian.dubuisson,laurent.sarry}@u-clermont1.fr
[2] Department of Medical Imaging, Notre-Dame Hospital, CHUM, Montreal, Canada
claude.kauffmann@gmail.com
[3] Department of Cardiology, Gabriel Montpied Hospital, 63001 Clermont-Ferrand, France
pmotreff@chu-clermontferrand.fr

Abstract. The aim of this study is to automatically assess reendothelialization of stents at an accuracy of down to a few microns by analyzing endovascular optical coherence tomography (OCT) sequences. Vessel wall and struts are automatically detected and complete distance map is then computed from sparse distances measured between wall and struts by thin-plate spline (TPS) interpolation. A reendothelialization score is mapped onto the geometry of the coronary artery segment. Accuracy and robustness are increased by taking into account the inhomogeneity of datapoints and integrating in the same framework orthogonalized forward selection of support points, optimal selection of regularization parameters by generalized cross-validation (GCV) and rejection of detection outliers. The comparison against manual expert measurements for a phantom study and 12 in vivo stents demonstrates no significant discordance with variability of the order of the strut thickness.

1 Introduction

Optical coherence tomography (OCT) is an intravascular imaging technique that provides the high-image resolution (<20μm) capable of assessing stented arterial segments in vivo. OCT can accurately differentiate the most superficial layers of the arterial wall as well as the stent struts and the vascular tissue surrounding them [1]. OCT could become a reference tool for assessing appropriate healing of stented coronary segments and comparing various types of stents, and could therefore guide optimal antiplatelet therapy to prevent late stent thrombosis.

Recent studies have confirmed the accuracy of OCT for quantifying neointimal hyperplasia and covered-strut rates with different stents [2]. Current methods of quantification are accurate and offer good inter- and intra-observer reproducibility, but they are still manual and the time-intensive image analysis step remains a major limitation. An automatic algorithm for detecting stent reendothelialization, i.e. thickness of neointimal cells covering the stent, has already been studied from volumetric optical coherence, but only on an *in vitro* model [3]. The blood vessel mimic did not reproduce the artifacts and difficulties with human OCT acquisition (move off center or angulation of the probe, coronary artery and cardiac movements, suboptimal saline flush) or stent malapposition. A semi-automatic method was developed with good

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 475–482, 2009.

results, but only for quantification of lumen, stent and neointimal areas, without strut coverage assessment [4]. We proposed to develop an automatic and supervised detection algorithm for lumen and struts (metallic echoes of the stent wire), making it possible to calculate prognostic OCT parameters used in clinical practice. We compared the algorithm results to conventional measures of neointimal hyperplasia and strut coverage on a dozen of human coronary stents analyzed by OCT 6 months after implantation.

2 Feature Detection

2.1 Wall Detection

The first part deals with the delineation of vessel bed edges. Depending on the amount of reendothelialization, those edges correspond to the interface of blood with metal (Fig. 1a) and/or with endothelium (Fig. 1b).

Vessel bed edge segmentation is obtained in a three-step process. At step 1, an automatic threshold level is computed on the gray-scale histogram of the 3D OCT image set using the Otsu method [5]. This threshold level is then applied to all cross-sectional OCT images in order to obtain a binary OCT image set. This binary image set is used at step 2 to find a rough approximation of the wall border. For this, each OCT section is run through a morphological segmentation process:

$$M_i = \left(H(T_i \circ K_1)\right) \cdot K_2 \,, \tag{1}$$

where T_i is the i^{th} thresholded OCT slice image, H is a hole-filling function, and K_1, K_2 are the structuring elements used by the function processing morphological opening and closing, respectively. The structuring elements K_1 and K_2 are binary discs with a given radius of $r_1 = 2$ and $r_2 = 9$ pixels respectively. The resulting morphological image M_i is a clean binary image of the vessel wall, where noise speckle, struts and artifacts (such as the guide at the image center) have been removed. The wall border can be sampled by profile analysis from lines drawn from the image center to the border at different angular positions. The detected wall position is used to initialize an active contour near to the real edges of the vessel bed. At step 3, the vessel bed edges are precisely segmented (Fig. 1) via a segmentation process based on the active contour model [6].

In case of strong malapposition stemming from the presence of thrombus or wall dissection, the exact wall boundary is sometimes difficult to recover (Fig. 1). Fortunately, the detected border is almost always irregular and an alert can be automatically sent to the user to control the detection result and correct it if necessary.

In order to cut the time needed to run the qualitative clinical validation of the vessel wall segmentation results, we perform an automatic shape analysis. As the vessel contour is normally expected to be circular or elliptic, an ellipse is fitted to the segmented vessel wall contours on each OCT image [7]. The Mean Square Error e between the vessel wall contour and the fitted ellipse is computed and used to trigger an alert for user validation if required. A simple alert threshold $e > T_A$ was used to indicate to the user which contours should be qualitatively validated and corrected if needed. A typical vessel wall contour that triggers an alert on is presented on Fig. 1.

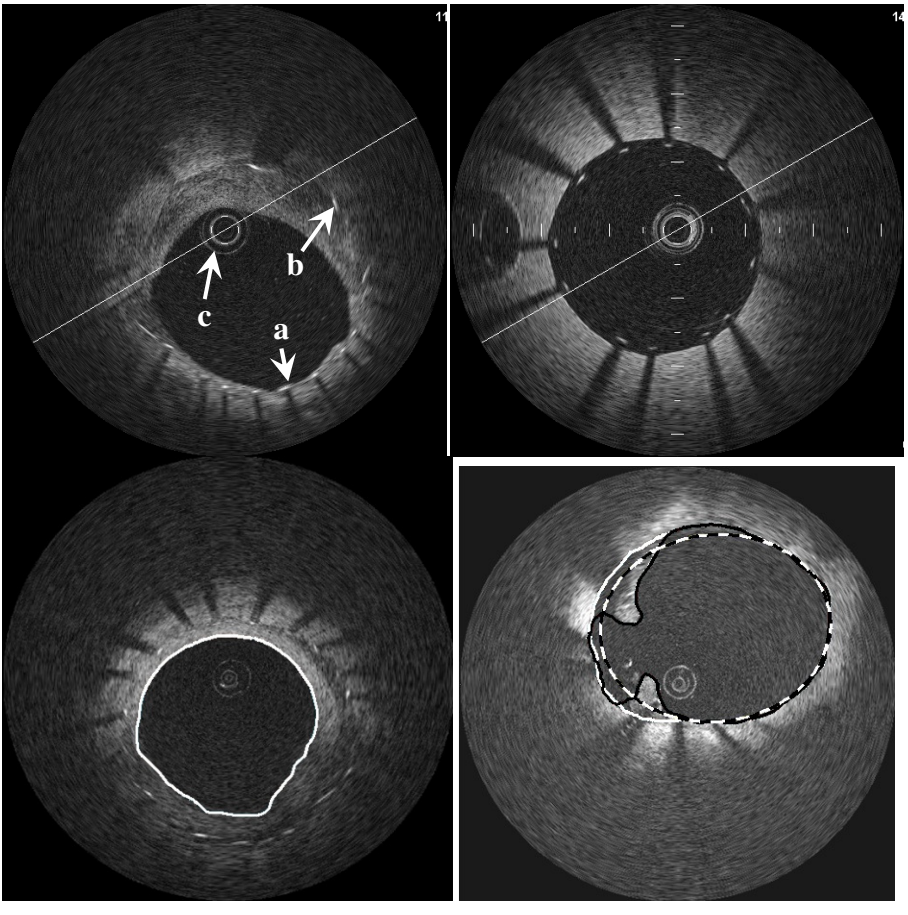


Fig. 1. Example of an OCT slice with inhomogeneous coverage (*top left*): some struts are in direct contact with blood (a), whereas others are completely covered by neointimal hyperplasia (b). The guide of the OCT probe is visible at the image center (c). Struts seem to be floating in the lumen just after stent deployment (*top right*). Examples of a good automatic segmentation result (*bottom left*), and one that fails (*bottom right, black line*), compared to a fitting ellipse (*dotted line*) to raise an alert for user correction (*white line*).

2.2 Strut Detection

The strut segmentation process presented in this section is performed in a slice-by-slice process. In order to simplify the strut segmentation algorithms, each OCT slice image is reformatted and interpolated in polar coordinates (θ, ρ).

The appearance of a strut in an OCT slice image is not unique and constant. In the best case, it appears as a high contrast speckle with a variable length depending on the orientation of the OCT probe with regard to the stent. When the contrast of the vessel wall is high, the strut produces a radial shadow zone that can be efficiently used to infer the presence or not of a strut. It is possible to enhance vertical lines produced by

the borders of the shadow zone by computing g_θ, the derivative with respect to θ. Secondly, we compute the cumulative of g_θ along the ρ-axis for each angle θ. Then, we search all specific increasing monotonic signal transitions, starting at a local minimum below zero and ending at a local maximum greater than zero. Two distinct metrics are defined to detect a shadow zone. The first is the peak-to-peak amplitude of the transition signal f_a in (2), and the second represents the symmetry between the absolute values of L_{max} and L_{min} respectively, noted f_b in (2):

$$f_a(i) = L_{max}(i) - L_{min}(i) \text{ and } f_b(i) = L_{max}(i) - |L_{min}(i)|\,, \tag{2}$$

where i stands for the i^{th} signal transition, $L_{max}(i)$ always > 0 and $L_{min}(i)$ always < 0.

A normalized metric of symmetry, $S(i)$ based on $f_a(i)$ and $f_b(i)$ is used in (3), where $S(i) = 1$represents a perfect symmetry and $S(i) = 0$is a perfect asymmetry:

$$S(i) = 1 - f_b(i)/f_a(i)\,. \tag{3}$$

A shadow zone is then detected if $f_a(i) > T_1$ and $S(i) > T_2$; T_1 and T_2 are threshold values, empirically set with regard to the values of $f_a(i)$ and $S(i)$, that are generated by noise and artifacts, when no shadow zone exists. The final step is to detect the exact position of the strut inside the shadow zone. For this, we analyze the profiles for each angular position defined inside the shadow zone and search the strut position by a priori knowledge of high gray level values.

But sometimes, the strut appears without shadow zone because the vessel wall is not highlighted. For this reason, we use a second pass algorithm to detect specific signal signature produced by strut without shadow zones. In order to enhance the strut detection in this case, the detection probability is strengthen when the strut is detected in successive angular positions so that a large strut becomes a higher probability as a small one, which can be a signal artifact.

3 Reendothelialization Interpolation

At this stage, the distance d_i to the vessel wall is only known at locations (θ_i, z_i) of the p detected struts. Due to detection uncertainties, some of the struts are missed, while others are outliers, i.e. their locations are totally abnormal. These outliers are really problematic because they can be falsely interpreted as under-deployed struts. Thus, it has to be possible to interpolate missing information and reject outliers at the same time.

A thin-plate spline (TPS) model was chosen in order to simulate the rigidity of the stent: it is made up of an affine part and a sum of functions $U(r) = r^2 \log(r)$,, bounded and asymptotically flat, centered at the struts (θ_i, z_i):

$$f(\theta, z) = a_1 + a_\theta \theta + a_z z + \sum_{i=1}^{p} w_i U(\|(\theta_i, z_i) - (\theta, z)\|)\,. \tag{4}$$

The regularized solution for (4) comes from the knowledge of the real-valued distance d_i at locations (θ_i, z_i) and is expressed in terms of a linear system:

$$\mathbf{Hw'}=\mathbf{H}\begin{bmatrix}\mathbf{a}\\ \mathbf{w}\end{bmatrix}=\begin{bmatrix}\mathbf{0} & \mathbf{P}^{\mathbf{T}}\\ \mathbf{P} & \mathbf{K}+\lambda\mathbf{I}_{\mathbf{p}}\end{bmatrix}\begin{bmatrix}\mathbf{a}\\ \mathbf{w}\end{bmatrix}=\begin{bmatrix}\mathbf{o}\\ \mathbf{d}\end{bmatrix}=\mathbf{y}, \tag{5}$$

where $K_{ij} = U(\|(\theta_i, z_i) - (\theta, z)\|)$,, the i^{th} row of $\mathbf{P}$ is $(1\ x_i\ y_i)$, $\mathbf{O}$ is a 3x3 null matrix, $\mathbf{o}$ is a 3x1 null column vector, $\mathbf{W}$ and $\mathbf{d}$ are column vectors formed from w_i and d_i, respectively, and $\mathbf{a}$ is the column vector containing elements a_1, a_θ and a_z.

In order to speed up the resolution process, orthogonalized forward selection is used. It consists in selecting a subset of $\boldsymbol{m}$ samples from the initial data, and the full design matrix $\mathbf{H}$ becomes $\mathbf{H}_m$, *i.e.* a $(p+3)\times(m+3)$ matrix where the last $\boldsymbol{m}$ columns of which correspond to the $\boldsymbol{m}$ support points [8].

For simplification, the affine and radial basis parts are estimated separately: first the coefficients for the affine part are estimated in a least-square sense:

$$\mathbf{a} = (\mathbf{P}^{\mathbf{T}}\mathbf{P})^{-1}\mathbf{P}^{\mathrm{T}}\mathbf{d}. \tag{6}$$

Then, the estimation for the radial function weights is performed on the residuals $\mathbf{r} = \mathbf{d} - \mathbf{Pa}$.. This is equivalent to ridge regression [9], *i.e.* minimization of first order regularization energy $E_r$equal to SSE plus λ times $\mathbf{w}_m^{\mathrm{T}}\mathbf{w}_m$; the weights $\mathbf{w}_m$ associated to the $\boldsymbol{m}$ support points are given by:

$$\mathbf{w}_m = (\mathbf{H}_m^{\mathrm{T}}\mathbf{H}_m + \lambda\mathbf{H}_m)^{-1}\mathbf{H}_m^{\mathrm{T}}\mathbf{r} \tag{7}$$

The choice of the $\boldsymbol{m}$ best centers represents a computational burden unless forward selection is used. This consists in picking the centers one after the other from among the remaining set until the decrease in a model selection criterion (MSC) is sufficiently low. At iteration $\boldsymbol{m}$, the design matrix $\mathbf{H}_{m-1}$is augmented by $\mathbf{h}_m$, the column of the full design matrix corresponding to the point that minimizes E_r.

It has been proven by Chen *et al.* [10] that Gram-Schmidt orthogonalization of the design matrix $\mathbf{H}_m$ in $\widetilde{\mathbf{H}}_m$, so that $\mathbf{H}_m = \widetilde{\mathbf{H}}_m\mathbf{U}_m$ with $\mathbf{U}_m$ an upper triangular, is likely to significantly speed up forward selection. Instead of choosing $\mathbf{h}_m$ among the columns of $\mathbf{H}$, its regularized version $\tilde{\mathbf{h}}_m$ is chosen from $\widetilde{\mathbf{H}}_m$ iteratively updated in the following way:

$$\widetilde{\mathbf{H}}_m = \widetilde{\mathbf{H}}_{m-1} - \frac{\tilde{\mathbf{h}}_m\tilde{\mathbf{h}}_m^{\mathrm{T}}\widetilde{\mathbf{H}}_{m-1}}{\tilde{\mathbf{h}}_m^{\mathrm{T}}\tilde{\mathbf{h}}_m} \text{ with } \widetilde{\mathbf{H}}_0 = \mathbf{H}, \tag{8}$$

and minimization of E_r means that $\tilde{\mathbf{h}}_j$ must maximize the gap in energy between two selections:

$$\Delta E_{rm}(j) = \frac{\left(\mathbf{r}^{\mathrm{T}}\tilde{\mathbf{h}}_j\right)^2}{\lambda + \tilde{\mathbf{h}}_j^{\mathrm{T}}\tilde{\mathbf{h}}_j}. \tag{9}$$

The transition matrix $\mathbf{U}_m$ itself is updated by:

$$\mathbf{U}_m = \begin{bmatrix}\mathbf{U}_{m-1} & \left(\widetilde{\mathbf{H}}_{m-1}^{\mathrm{T}}\widetilde{\mathbf{H}}_{m-1}\right)^{-1}\widetilde{\mathbf{H}}_{m-1}^{\mathrm{T}}\mathbf{h}_m\\ \mathbf{0}_{m-1}^{\mathrm{T}} & 1\end{bmatrix} \text{ with } \widetilde{\mathbf{U}}_0 = 1, \tag{10}$$

and makes it possible to recover final weights:

$$\mathbf{w}_m = \mathbf{U}_m^{-1}\widetilde{\mathbf{w}}_m \tag{11}$$

from regularized weights:

$$\tilde{\mathbf{w}}_m = \frac{\mathbf{r}^{\mathrm{T}}\tilde{\mathbf{h}}_m}{\lambda + \tilde{\mathbf{h}}_m^{\mathrm{T}}\tilde{\mathbf{h}}_m}, \tag{12}$$

simply by back-substitution.

The MSC used as stopping criterion for the previous iterative algorithm is generalized cross-validation (GCV) [11,12] as it is considered to avoid data overfit. In [9], it is expressed as:

$$GCV_m = \frac{\mathbf{r}^{\mathrm{T}}\mathbf{P}_m^2\mathbf{r}}{\left(\mathrm{trace}(\mathbf{P}_m)\right)^2} \text{ where } \mathbf{P}_m = \mathbf{I}_p - \tilde{\mathbf{H}}_m\left(\tilde{\mathbf{H}}_m^{\mathrm{T}}\tilde{\mathbf{H}}_m + \lambda\tilde{\mathbf{H}}_m\right)^{-1}\tilde{\mathbf{H}}_m^{\mathrm{T}}, \tag{13}$$

is the matrix projecting the data $\mathbf{r}$ perpendicular to the space spanned by the model.

Orr has shown that the estimate of regularization parameter λ can be inserted in the algorithm, and that its value can be refined at every iteration just after selection of the new center, by means of a re-estimation formula coming from the minimization of GCV_m: its complete derivation can be found in [9].

4 3D Visualization of the Reendothelialization Score

The wall edges detected in every slice are used to build a mesh of the vessel lumen by using the Marching Cubes algorithm [13]. Then each vertex is assigned a color corresponding to the reendothelialization thickness at its location in cylindrical coordinates (θ, z).

The following color look-up table is used: green for well-apposed to the vessel wall with apparent neointimal coverage; red for well-apposed but without coverage (absence of any visible structure between lumen and vessel on OCT confirmed by a strong echo at the strut), and black for malapposed without coverage (malapposition being defined as a distance higher than the width of the strut). Figure 2 gives an example of a stent with inhomogeneous coverage.

5 Automatic Estimation versus Manual Measurement

First, a phantom-based experiment consisting in deploying a stent into a urinary catheter and analyzing its position using OCT gave an error of 2 µm and a variability of 6.72 µm with regard to the theoretical thickness of the metallic wire.

Second, *in vivo* cross-sections taken every millimeter from 12 OCT pull-back recordings over stented segments were manually analyzed for a total of 2735 individual struts. Lumen Area (LA, in mm²) and Stent Area (SA, in mm²) were measured on each image to compute the percentage of NeoIntimal Hyperplasia Area as NIHA = 100(SA-LA)/SA. Neointimal hyperplasia thickness was measured manually by a cardiologist for each strut using the offline software delivered with the OCT console (LightLab Imaging, Massachusetts, USA). It was defined as the closest distance between one strut and the lumen contour, but was difficult to be accurately measured manually and therefore cannot be considered as a perfect gold standard.

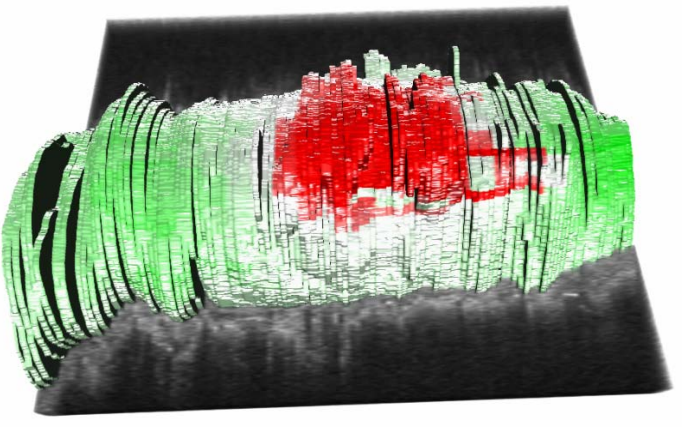


Fig. 2. 3D representation of reendothelialization score on the vessel lumen. For this stent, two scores are represented: well-apposed with coverage (green) and well-apposed without coverage (red). Relative displacements between slices are due to OCT probe movement inside lumen due to cardiac cycle.

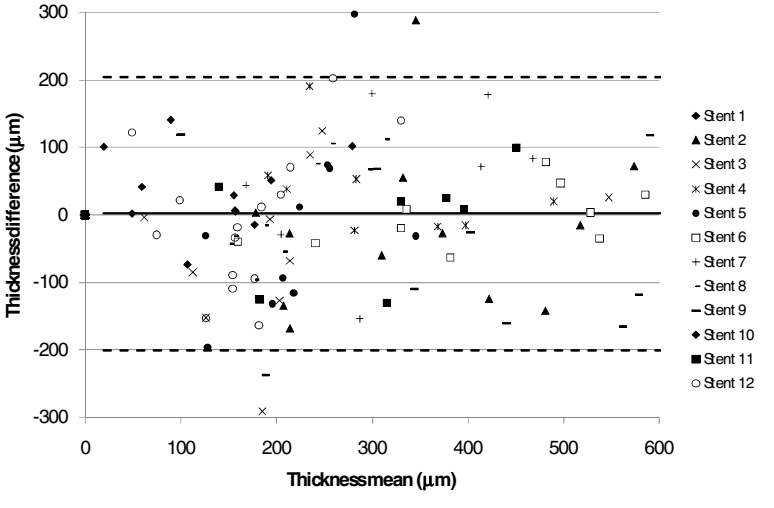


Fig. 3. Bland-Altman plot of manual versus automatic thickness measurements expressed in µm. The bias 95% confidence interval is equal to]-1.68;5.91[and standard deviation is equal to 101.2 µm. For the sake of clarity, one point out of 25 is drawn.

NIHA automatic rate was overestimated by 1.13±1.45% on average with regard to manual measurement and ranged from 1.53 to 50.18%. Figure 3 illustrates that the two methods are not significantly discordant (risk 5%) and therefore could be exchanged since agreement bounds correspond approximately to the admissible clinical limit, i.e. the order of magnitude of the metallic wire thickness (150 µm). As stated above, most of the variability might also be explained by manual data; it is confirmed

by the high precision of NIHA estimation that does not depend on the assumption of closest distance.

Processing time decreases from 4 hours for manual analysis (1 slice out of 15 being analyzed) down to a maximum of 10 minutes in the worse case of a sequence that required user interaction to correct 8 lumen contours raising alerts.

Future prospects concern the correction of OCT movement artifacts in order to obtain a smoother score representation and the recovery of the true 3D geometry by estimating the probe pull-back path from angiography.

Acknowledgments. The authors are grateful for the partnership with Object Research Systems Company that supplied visualization platform ORS Visual and SDK and for their valuable assistance for developments in computer graphics.

References

1. Pinto, T.L., Waksman, R.: Clinical applications of optical coherence tomography. J. Interv. Cardiol. 19, 566–573 (2006)
2. Chen, B.X., Ma, F.Y., Luo, W., Ruan, J.H., Xie, W.L., Zhao, X.Z., Sun, S.H., Guo, X.M., Wang, F., Tian, T., Chu, X.W.: Neointimal coverage of bare-metal and sirolimus-eluting stents evaluated with optical coherence tomography. Heart 94, 566–570 (2008)
3. Bonnema, G.T., Cardinal, K.O., McNally, J.B., Williams, S.K., Barton, J.K.: Assessment of blood vessel mimics with optical coherence tomography. J. Biomed. Opt. 12, 24018 (2007)
4. Tanimoto, S., Rodriguez-Granillo, G., Barlis, P., de Winter, S., Bruining, N., Hamers, R., Knappen, M., Verheye, S., Serruys, P.W., Regar, E.: A novel approach for quantitative analysis of intracoronary optical coherence tomography: high inter-observer agreement with computer-assisted contour detection. Catheter Cardiovasc. Interv. 72, 228–235 (2008)
5. Otsu, N.: A threshold selection method from gray-scale histogram. IEEE Trans. Systems, Man, and Cybernetics 8, 62–66 (1978)
6. Kauffmann, C., Godbout, B., de Guise, J.A.: Simplified active contour model applied to bone structure segmentation in digital radiographs. In: Medical Imaging, SPIE International Symposium, pp. 663–672 (1998)
7. Fitzgibbon, A., Pilu, M., Fisher, R.B.: Direct least square fitting of ellipses. IEEE Trans. Patt. Anal. Mach. Intell. 21, 476–480 (1999)
8. Donato, G., Belongie, S.: Approximate thin plate spline mappings. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2352, pp. 21–31. Springer, Heidelberg (2002)
9. Orr, M.J.L.: Regularisation in the selection of radial basis function centres. Neural Computation 7, 606–623 (1995)
10. Chen, S., Cowan, C.F.N., Grant, P.M.: Orthogonal least squares learning for radial basis function network. IEEE Trans. Neural Networks 2, 302–309 (1991)
11. Wahba, G.: Spline Models for Observational Data. In: CBSM-NSF Regional Conf. Ser. Appl. Math., vol. 59. SIAM, Philadelphia (1990)
12. Bartoli, A.: Maximizing the predictivity of smooth deformable image warps through cross-validation. J. Math. Imaging Vis. 31, 133–145 (2008)
13. Lorensen, W.E., Cline, H.E.: Marching Cubes: a high resolution 3D surface reconstruction algorithm. Computer Graphics 21, 163–169 (1987)

Optical Biopsy Mapping for Minimally Invasive Cancer Screening

Peter Mountney[1,2], Stamatia Giannarou[2], Daniel Elson[2], and Guang-Zhong Yang[1,2]

[1] Department of Computing
[2] Institute of Biomedical Engineering Imperial College, London SW7 2BZ, UK
{peter.mountney,stamatia.giannarou03,ds.elson, g.z.yang}@imperial.ac.uk

Abstract. The quest for providing tissue characterization and functional mapping during minimally invasive surgery (MIS) has motivated the development of new surgical tools that extend the current functional capabilities of MIS. Miniaturized optical probes can be inserted into the instrument channel of standard endoscopes to reveal tissue cellular and subcellular microstructures, allowing excision-free optical biopsy. One of the limitations of such a point based imaging and tissue characterization technique is the difficulty of tracking probed sites in vivo. This prohibits large area surveillance and integrated functional mapping. The purpose of this paper is to present an image-based tracking framework by combining a semi model-based instrument tracking method with vision-based simultaneous localization and mapping. This allows the mapping of all spatio-temporally tracked biopsy sites, which can then be re-projected back onto the endoscopic video to provide a live augmented view in vivo, thus facilitating re-targeting and serial examination of potential lesions. The proposed method has been validated on phantom data with known ground truth and the accuracy derived demonstrates the strength and clinical value of the technique. The method facilitates a move from the current point based optical biopsy towards large area multi-scale image integration in a routine clinical environment.

1 Introduction

With recent advances in biophotonics and surgical instrumentation, there is an increasing demand to bring cellular and molecular imaging modalities to an *in vivo*, *in situ* setting to allow for real-time tissue characterization, functional assessment and intra-operative guidance. Miniaturization of the confocal laser scanning microscope, for example, has led to imaging probes that can be inserted into the instrument channel of a standard endoscope to visualize cellular and subcellular microstructures to provide 'optical biopsy' without excision of tissue. Following the application of a contrast agent, this can allow for the detection of colorectal adenomas, disruption in the pit pattern of the colon, angiogenesis, and neoplasia in Barrett's esophagus [1]. It has also been used without a contrast agent to detect malignant disruption of the bronchial basement membrane using elastin autofluorescence [2]. Other techniques that enable microscopic detection and characterization of tissue include Optical Coherence

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 483–490, 2009.

Tomography (OCT), two photon excited fluorescence and high magnification endoscopy [3]. There have also been successful clinical trials of techniques that acquire detailed spectroscopic information for cancer detection, for example using the time- or wavelength-resolved fluorescence or Raman properties.

For *in vivo* applications, all of these techniques suffer from the limitation of only providing a small, localized probe region whilst the organs of interest may require a large surface area to be surveyed. Technically, the main difficulty of tracking the optical biopsy sites is that these probes leave no marks on the tissue. Furthermore, the optical biopsy sites move in and out of the view in a standard endoscope image as the examination progresses and may deform as a result of respiration or tissue-instrument interaction. Current approaches to long-term tissue-instrument tracking assume the use of rigid laparoscopes and availability of optical markers [4]. Structure from motion has been used to reconstruct 3D tissue models, but it suffers from drift and does not work well when revisiting biopsy sites [5]. For extending the effective field-of-view of the endoscopic image, image mosaicing [6] and dynamic view expansion [7] have been used to reconstruct enlarged field-of-views, although these techniques tend not to explicitly deal with motion parallax.

In practice, optical probes are typically introduced through the instrument channel while holding the endoscope stationary. Since the probe needs to be placed in contact with the tissue when the optical biopsy takes place, tracking the tip of the probe enables the localization of the biopsy site. To this end, it is necessary to take into account scale, rotation and illumination changes when tracking the tool. Current approaches to needle and surgical instrument tracking may be applicable [8, 9], but a combined approach by integrating probe tracking with a 3D probabilistic map built *in situ* using only white light endoscopic images with no additional fiducials can ensure robustness and practical clinical use. This work proposes an image-based tracking system based on SLAM (Simultaneous Localization and Mapping) for optical probes. This will allow for subsequent localization and contextual analysis of microstructures or guiding real tissue biopsy. The main contribution of this paper is to combine SLAM with probe tracking to create a 3D model of the tissue surface and spatio-temporally tracked optical biopsy sites. These biopsy sites are subsequently re-projected back onto the image plane to provide a live augmented view *in vivo*, thus facilitating re-targeting and serial examination. The proposed method has been

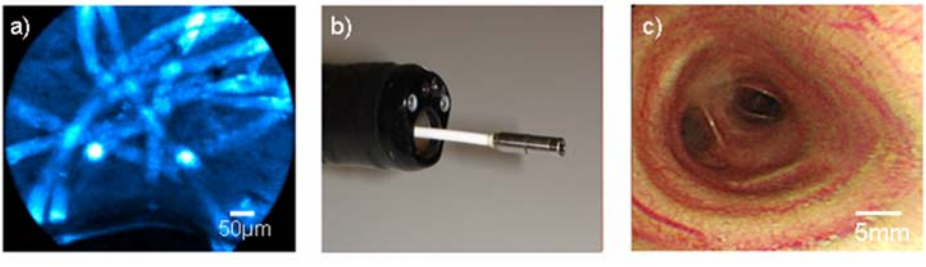


Fig. 1. (a) A typical microconfocal fluorescence image showing the microstructure of a sample, (**b)** the relative configuration of a confocal fluorescence probe when inserted through the instrument channel of a standard endoscope, and (**c)** a typical endoscopic white light image of the bronchus used for navigation

validated on phantom data with known ground truth. The method will facilitate a shift from the current point based optical biopsy towards multi-scale image integration in a routine clinical environment.

2 Methods

2.1 Probabilistic Mapping

The first step of the proposed tracking framework is to establish a probabilistic mapping of the environment. Previous work on SLAM based approaches has shown the ability to generate 3D tissue models and recover the relative pose of the endoscope [10]. A long-term map is generated making it more resilient to drift and error accumulation over time, and thus is well suited to returning to previously targeted areas. In this work, a vision based sequential approach has been used. This is based on an Extended Kalman Filter (EKF) framework with state vector x containing the position (c_x, c_y, c_z), orientation $(c_{q1}, c_{q2}, c_{q3}, c_{q4})$, translational velocity (v_x, v_y, v_z) and angular velocity $(\omega_x, \omega_y, \omega_z)$ of the endoscope. In addition, the state vector also stores 3D locations of salient features in the map (y_x, y_y, y_z). A constant velocity, constant angular velocity motion model is used to predict the endoscope's motion with Gaussian noise. Accompanying the state vector is the covariance matrix which stores the uncertainty of the endoscope and feature locations in 3D. In this sequential map building approach, new features are added to the map on the fly by feature matching constrained by epipolar geometry to estimate their 3D positions relative to the endoscopic camera.

2.2 Biopsy Site Estimation

The initial position of the biopsy site in the image plane is estimated through probe tracking. In this work, no marker was attached and no changes were made to the colour of the imaging probe. The technique exploits the fact that the camera is relatively static when the biopsy is taken. The segmentation of the tool is achieved by combining background subtraction and color segmentation in the HSV space. In this study, a simple background subtraction technique is used based on inter-frame difference. Foreground/background models are learnt and updated over time. The background model is initialized with the first frame of the video sequence. For the extraction of foreground objects, the current frame is subtracted from the background model and any significant difference is labeled as foreground. If no foreground object is identified, the current frame becomes the background model. On the saturation plane, the shaft of the probe is highlighted in dark grey on a bright background. Therefore, foreground pixels (Fig. 2 (a)) are used as seeds for region growing in the saturation color plane (Fig. 2 (b)) to segment the probe shaft as shown in Fig. 2(c).

In order to identify the tip of the probe, the centroid of the shaft is extracted. The tangentials of the shaft are detected at the global maxima of the Hough transform and the axis of the shaft is computed as the eigenvector corresponding to the smallest eigenvalue of the moment of inertia. The localization of the tip of the probe is performed with respect to a reference point located at the intersection of the shaft and the

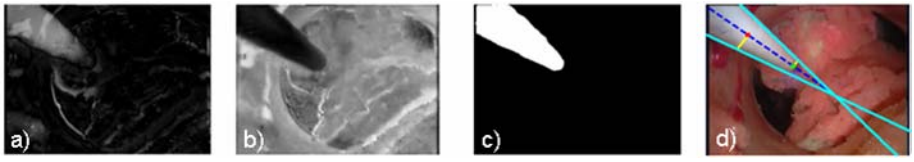


Fig. 2. Probe tracking and biopsy site estimation within the image plane; **(a)** background subtraction, (**b)** color saturation distribution within the image, (**c)** segmented tool regions, and (**d)** the model fitted tool (centroid -red dot, the reference point - green dot, vanishing tangential lines - cyan, radius at the center of mass and at the reference point – yellow).

distal tip. The 3D position of the tip is estimated using a semi-model based approach assuming rigidity and incorporating prior knowledge of the width of the probe and the relationship between the reference point and the tip. The position of the reference point and the orientation of the shaft are estimated in 3D and the prior model enables the localization of the tip of the probe b^c relative to the camera.

2.3 Global Biopsy Mapping

Following the steps in Section 2.2, the position of biopsy site b^c is estimated in the camera coordinate system. It is transformed into the world coordinate system using:

$$b^w = C^w b^c + c^w \tag{1}$$

where b^w is the biopsy site in the world coordinate system, C^w and c^w are the orientation and position of the camera in the global SLAM coordinate system. Although the 3D position of the biopsy site is now defined, this position is never directly observed or measured again. There are two reasons for this; the actual site on the tissue is usually occluded by the probe when the biopsy is taken, and there may not be any salient features at or around the biopsy site to be tracked. In this case, 2D tracking would fail. However, the strength of the proposed probabilistic map is that the position of the biopsy site can be updated without directly measuring it. This is made possible by the co-variance matrix which models the uncertainty of all the biopsy positions. The i^{th} biopsy site b_i^w is inserted into the state vector and the co-variance matrix P is updated. The co-variance matrix is updated with the partial derivatives $\partial b_i / \partial x_v$ of the biopsy site with respect to the camera position, as well as the measurement model $\partial b_i / \partial h_i$ and measurement noise R as shown in Eq (2).

$$b_i^w = \begin{pmatrix} x \\ y \\ z \end{pmatrix}, \qquad P = \begin{bmatrix} P_{xx} & P_{xy_1} & P_{xx}\frac{\partial b_i^T}{\partial x_v} \\ P_{y_1x} & P_{y_1y_1} & P_{y_1x}\frac{\partial b_i^T}{\partial x_v} \\ \frac{\partial b_i^w}{\partial x_v}P_{xx} & \frac{\partial b_i^w}{\partial x_v}P_{xy_1} & \frac{\partial b_i^w}{\partial x_v}P_{xx}\frac{\partial b_i^{wT}}{\partial x_v} + \frac{\partial b_i^w}{\partial h_i}R\frac{\partial b_i^{w\,T}}{\partial h_i} \end{bmatrix} \tag{2}$$

where x is the position of the endoscope and y_i is the i^{th} feature in the map.

The position and uncertainty of the biopsy sites are correlated to the camera position and the rest of the features in the map. Fig. 3 illustrates this sequential map building demonstrating how the camera, features and the biopsy sites are correlated and

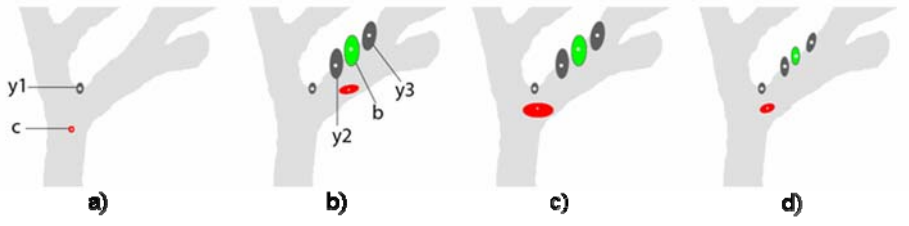


Fig. 3. (a-d) Schematic representation of sequential probabilistic mapping updates. The camera's position c is shown in red with the uncertainty represented by an ellipse, features y1, y2 and y3 represented in dark gray, the biopsy site b shown in green and the tissue shown in light gray. **(a)** c measures y1 with low uncertainty, **(b)** c is navigated to a new position with growing uncertainty. Features y2 and y3 are measured and biopsy b is taken. **(c)** c is navigated close to y1 and positional uncertainty increases. **(d)** Feature y1 is measured and the position estimate of c is improved. Resulting in an improved estimate of b, as it is correlated to c.

temporally updated. At the time when the biopsy site is observed, the uncertainty of the camera's position may be high, as illustrated in Fig. 3 (b) but the relative position of the biopsy site to surrounding features is well defined. Over time, the camera will re-measure these surrounding features in the map as in Fig. 3 (d) and the position estimation of the camera will improve, thus reducing the uncertainty. Therefore, the position estimation of the biopsy site will also improve as it is correlated to the position of the camera and will not drift away in the global map. To facilitate real-time examination, the biopsy sites $\{b_1^w \ldots b_i^w\}$ are visualized in this study by re-projecting the 3D points into the camera plane based on the intrinsic camera parameters and the estimated camera position from SLAM. This provides an augmented view of the biopsy sites for the operator.

2.4 Experimental Set-Up

The proposed approach has been validated on a silicon phantom of the airway coated with acrylic paint to provide realistic texture and internal reflections. Sponge cell structures were attached to the internal surface to enable optical biopsies to be taken using a confocal fluorescence endoscope system (Cellvizio, Mauna Kea Technologies, Paris). Validation was performed by measuring the accuracy of biopsy sites in the image space as the endoscope navigated through the phantom. The ground truth data used for comparison was collected using an optical tracking device (Northern Digital Inc, Ontario, Canada) and an experienced observer. To obtain the ground truth position of the camera, a rigid stereo laparoscope fitted with four optical markers was used. The position c^l and orientation C^l of the center of the left camera relative to the optical markers were acquired using standard hand-eye calibration [11]. This enabled the position of the camera to be calculated in the world coordinate system c^w and C^w .To obtain the ground truth of the 3D biopsy site positions, the experienced observer manually identified the sites on the stereo images at the time when the biopsy was taken. By using the camera's intrinsic and extrinsic parameters, the 3D position b^c of the biopsy site was obtained relative to the camera, and its position in

the world coordinate system b^w was determined as $b^w = C^w * b^c + c^w$. At each subsequent frame, the biopsy site b^w was projected into the ground truth camera position $x = x_o - fk_x(b_x^c / b_z^c)$ and $y = y_o - fk_y(b_y^c / b_z^c)$ where fk_x and fk_y are the focal length and x_o and y_o are the principal point. To validate the proposed probe tracking approach, the probe was mounted in a rigid sheath. This evaluation step was combined with manually defined image coordinates of the probe's location.

3 Results

The proposed algorithm was validated on a two minute long stereo laparoscopic video sequence consisting of navigation to four different areas, including six biopsies and re-targeting previously taken biopsies.

Quantitative analysis of the position of the biopsy sites in the image plane is shown in Table 1. The average visual angle error for the position of the biopsy sites ranges from 1.18° to 3.86°. Figs. 4 (d-e) show the estimated biopsy site position and ground truth position of site three over a short sequence before the site goes out of view. Accuracy of the biopsy position estimation is affected by the proximity of the camera to the site where close proximity leads to a magnification of the error. Fig. 5 illustrates the results of the augmented biopsy sites at different stages of the procedure where changes in illumination, scale and view point are experienced. Fig. 5 demonstrates the practical value and clinical relevance of the proposed method; the entire procedure is represented where six biopsies are taken and added to the global map, including the associated biopsy images of the sponge cell structures.

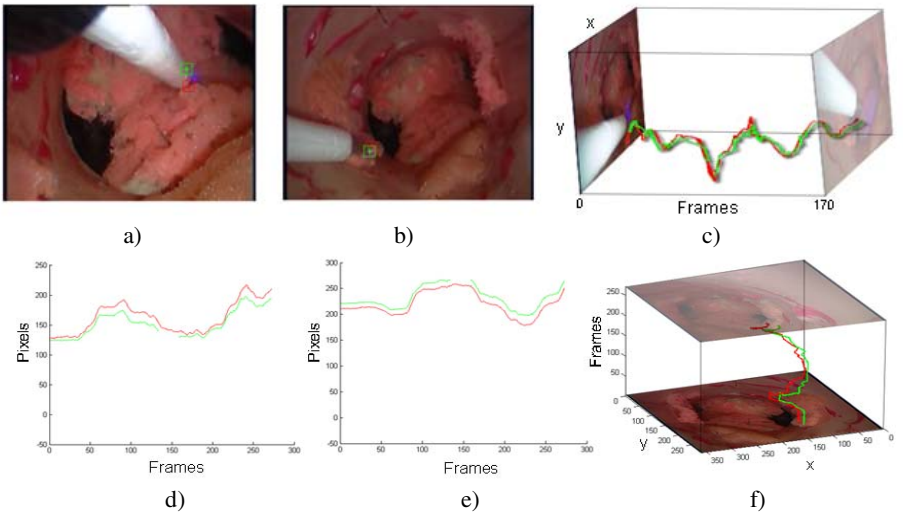


Fig. 4. (a-c) Probe tracking: Ground truth (red) and estimated (green) position of probe at **(a)** site six and **(b)** site three. **(c)** Ground truth (red) and tracked probe position (green) during navigation between biopsy sites. **(d-f)** Augmented biopsy site three: **(d-e)** the X and Y projected pixel error showing the site being tracked **(f)** the ground truth projected position (red) and the estimated position (green) for a short section of the procedure.

Table 1. Average error of biopsy site estimation and probe tracking for phantom experiments

Biopsy sites	Probe tracking		Augmented biopsy sites	
	Average visual angle error	Percent of FOV	Visual angle error	Percent of FOV
1	1.85°	4.89%	2.34°	5.37%
2	1.33°	3.59%	3.06°	7.58%
3	0.87°	2.29%	2.22°	5.59%
4	0.86°	2.23%	1.18°	2.99%
5	0.81°	1.75%	2.06°	4.61%
6	3.22°	8.88%	3.86°	10.09%

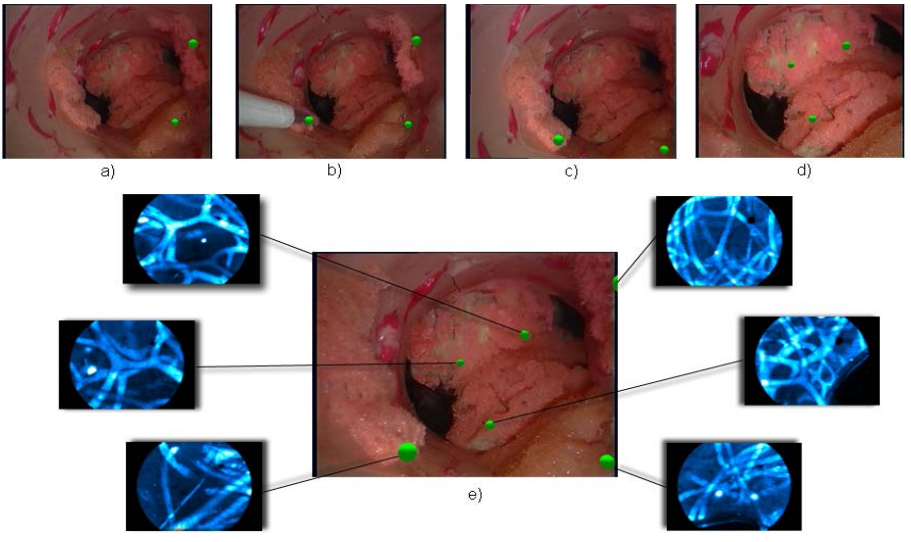


Fig. 5. (a-d) Biopsy site position (green spheres). The spheres are 2mm in diameter and appear in different sizes when they are projected onto the image under perspective projection; **(e)** shows the six biopsy sites with corresponding micro-confocal fluorescence endoscope images.

Detailed quantitative analysis of the probe tracking when the biopsies are taken is shown in Table 1. The tracking errors range from 0.81° to 3.22° of the visual angle and an example error distribution is illustrated in Fig. 4 (a-b). Quantitative analysis of the probe tracking on the whole sequence gave an average visual angle error of 2.87°. The sensitivity and specificity were 0.9706 and 0.9892, respectively. As expected, the accuracy deteriorates when the probe is introduced and removed from the scene as a part of the shaft is occluded, or when the probe is very close to the camera.

4 Conclusion

In this paper we have proposed a novel approach for microconfocal optical biopsy tracking which can be used to augment intra-operative navigation and retargeting of previously examined tissue regions. The system has been validated with a detailed

phantom experiment and we have demonstrated that this approach can accurately project the location of biopsy sites, thus enabling its practical clinical use. The proposed method requires no prior information of the tissue geometry and can operate consistently in a sparse feature environment. The proposed method is robust to small local deformation and rigid global motion. Modeling large scale nonlinear tissue deformation, however, is not trivial and will be addressed in future work.

Acknowledgments. We gratefully acknowledge support from the EPSRC and the Technology Strategy Board grants DT/F003064/1 and DT/E011101/1.

References

1. Meining, A., Bajbouj, M., von Delius, S., Prinz, C.: Confocal Laser Scanning Microscopy for in vivo Histopathology of the Gastrointestinal Tract. Arab Journal of Gastroenterology 8, 1–4 (2007)
2. Thiberville, L., Moreno-Swirc, S., Vercauteren, T., Peltier, E., Cavé, C., Bourg Heckly, G.: In Vivo Imaging of the Bronchial Wall Microstructure Using Fibered Confocal Fluorescence Microscopy. American Journal of Respiratory and Critical Care Medicine 175, 22–31 (2007)
3. Van Dam, J.: Novel methods of enhanced endoscopic imaging. GUT 52(4) (2003)
4. Wengert, C., Cattin, P.C., Duff, J.M., Székely, G.: Markerless Endoscopic Registration and Referencing. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 816–823. Springer, Heidelberg (2006)
5. Wu, C.-H., Sun, Y.-N., Chen, Y.-C., Chang, C.-C.: Endoscopic Feature Tracking and Scale-Invariant Estimation of Soft-Tissue Structures. IEICE Transactions on Information and Systems 2, 351–360 (2008)
6. Atasoy, S., Noonan, D.P., Benhimane, S., Navab, N., Yang, G.-Z.: A global approach for automatic fibroscopic video mosaicing in minimally invasive diagnosis. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 850–857. Springer, Heidelberg (2008)
7. Lerotic, M., Chung, A.J., Clark, J., Valibeik, S., Yang, G.-Z.: Dynamic View Expansion for Enhanced Navigation in Natural Orifice Transluminal Endoscopic Surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 467–475. Springer, Heidelberg (2008)
8. Wengert, C., Bossard, L., Häberling, A., Baur, C., Székely, G., Cattin, P.C.: Endoscopic Navigation for Minimally Invasive Suturing. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 620–627. Springer, Heidelberg (2007)
9. Krupa, A., Gangloff, J., Doignon, C., de Mathelin, M.F., Morel, G., Leroy, J., Soler, L., Marescaux, J.: Autonomous 3-D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing. IEEE Transactions on Robotics and Automation 19, 842–853 (2003)
10. Mountney, P., Stoyanov, D., Davison, A.J., Yang, G.-Z.: Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 347–354. Springer, Heidelberg (2007)
11. Tsai, R., Lenz, R.: Real Time Versatile Robotic Hand/Eye Calibration using 3D Machine Vision. In: Proc. ICRA 1988, pp. 554–561 (1988)

Biopsy Site Re-localisation Based on the Computation of Epipolar Lines from Two Previous Endoscopic Images

Baptiste Allain[1], Mingxing Hu[1], Laurence B. Lovat[2], Richard Cook[3], Sebastien Ourselin[1], and David Hawkes[1]

[1] Centre for Medical Image Computing, University College London UCL
b.allain@ucl.ac.uk
[2] Department of Gastroenterology, University College London Hospitals UCLH NHS Foundation Trust and National Medical Laser Centre, Division of Surgical Sciences, UCL
[3] Dental Institute, Department of Biomaterials, King's College London, KCL

Abstract. Tracking biopsy sites in endoscopic images can be useful to provide a visual aid for the guidance of surgical tools, for example when endoscopic guided biopsy is required. A new method for re-localisation of these sites is presented in this paper. It makes use of epipolar geometry properties between three images of the same site observed from different viewpoints with an endoscope. Two epipolar lines are derived from the two first images in the third image where the site needs to be re-localised. Their intersection corresponds to the location of the biopsy site. This method was tested with gastroscopic data from 2 patients with 9 series of three images of the oesophagus. The re-localisation error was estimated at less than 1.5 millimetres by a clinical endoscopist, which is sufficient for most clinical endoscopic applications.

1 Introduction

Endoscopy is a procedure for real-time video acquisitions of the interior of hollow organs using flexible or rigid instruments and facilitates taking biopsies. The main problem for many endoscopists is the detection of macroscopically invisible biopsy sites and re-localisation during the same or future examinations. CT and electromagnetic based systems for tracking the location of bronchoscopes have been described. These allow accurate targeting of biopsies in the relatively rigid bronchial tree [1, 2]. A precision of re-localisation of 1.58 mm has been reported. Some examinations like gastroscopy are based, however, entirely on visual images. In Barrett's oesophagus, precancerous lesions may not be visible under white light endoscopy but may be detected histologically. It is important to be able to re-localise these lesions, particularly after taking a virtual 'optical' biopsy or *in vivo* histology based solely on the information within the endoscopy view. For several 'optical' biopsy techniques, one solution is to pass a miniprobe via the endoscope working channel and place it in contact with the tissue to scan the surface. The scan helps decide whether a tissue specimen needs to be taken. The forceps then need to be re-localised where the miniprobe was positioned. Unfortunately, the optical measurement and the excised tissue may not match, which can make the biopsy irrelevant if negative [3, 4]. Thus, the detected biopsy site may need to be re-localised precisely in subsequent video images. Methods for object

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 491–498, 2009.

localisation in video images have been recently addressed in the field of robot-assisted surgery [5, 6], but have focused only on tracking the instrument tip. The scope of the biopsy site re-localisation focuses more on the projection of a point location from an image to another and requires the computation of the spatial transformations (rotations and translations) between endoscopic images. It has been demonstrated in [7] that the recovery of the epipolar geometry [8] between endoscopic images can be used to compute these spatial transformations.

The method presented in this paper proposes a solution to the re-localisation problem using robust feature tracking and the constraints of epipolar geometry for applications in gastroenterology. A point's location is computed in the target image as the intersection of two epipolar lines derived from the location of corresponding biopsy sites in pairs of previously acquired endoscopic images. The epipolar geometry is recovered with a robust technique. This method presents several advantages. (i) First, it reduces the computational burden since it is based only on the epipolar lines consideration. There is no need to directly compute the rotations and the translations between the images. (ii) Secondly, this method has the potential to be a real-time guidance system based on a limited amount of information. Biospy sites are tracked directly in the video images and not in relation to a pre-operative image. (iii) Finally, this method could be extended to any type of intra-operative optical imaging.

2 Method

2.1 Re-localisation Principle

If a biopsy site location is known in an endoscopic image $\mathbf{I_1}$, it can be projected onto another image $\mathbf{I_2}$ for its re-localisation. This projection can be determined with the epipolar geometry formed with the two images $\mathbf{I_1}$ and $\mathbf{I_2}$. The epipolar geometry between $\mathbf{I_1}$ and $\mathbf{I_2}$ can be described algebraically using the fundamental matrix $\mathbf{F_{12}}$ as:

$$\mathbf{F}_{12} = \begin{pmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{pmatrix} = \mathbf{K}_2^{-T}[\mathbf{t}_{12}]\mathbf{R}_{12}\mathbf{K}_1^{-1} = \mathbf{K}_2^{-T}\begin{bmatrix} 0 & -t_3 & t_2 \\ t_3 & 0 & -t_1 \\ -t_2 & t_1 & 0 \end{bmatrix}\mathbf{R}_{12}\mathbf{K}_1^{-1}. \quad (1)$$

$\mathbf{K_1}$ and $\mathbf{K_2}$ are the intrinsic matrices of the camera at first and second positions defined from the focal length, the position of the image centre, and the scaling from the 3D-space to the image. Once $\mathbf{F_{12}}$, $\mathbf{K_1}$ and $\mathbf{K_2}$ are known, it is possible to determine the camera motions, rotation $\mathbf{R_{12}}$ and translation $\mathbf{t_{12}}$, with further computations.

During endoscopic procedures, a biopsy site can be seen from various viewpoints with an endoscopic camera (Fig. 1). Let $\mathbf{I_1}$ and $\mathbf{I_2}$ be two images where the biopsy site location is visible and $\mathbf{I_3}$ be a third image where it needs to be re-localised. Let $\mathbf{P}$ be the biopsy site location in the 3D space, and $\mathbf{p_{T1}}$ and $\mathbf{p_{T2}}$ be the locations of the biopsy site in images $\mathbf{I_1}$ and $\mathbf{I_2}$. The fundamental matrices $\mathbf{F_{13}}$ and $\mathbf{F_{23}}$ can be computed between respectively images $\mathbf{I_1}$ and $\mathbf{I_3}$ and images $\mathbf{I_2}$ and $\mathbf{I_3}$. The axes formed with camera centre 1 and camera centre 3, and camera centre 2 and camera centre 3, have an intersection with the image plane $\mathbf{I_3}$, which is called the epipole. Let $\mathbf{e^{13}}$ and $\mathbf{e^{23}}$ be the two epipoles of this configuration. Thus, $\mathbf{F_{13}p_{T1}}$ is a vector and defines the epipolar line l_1, which passes through the projection of $\mathbf{p_{T1}}$ onto $\mathbf{I_3}$ and through $\mathbf{e^{13}}$. The

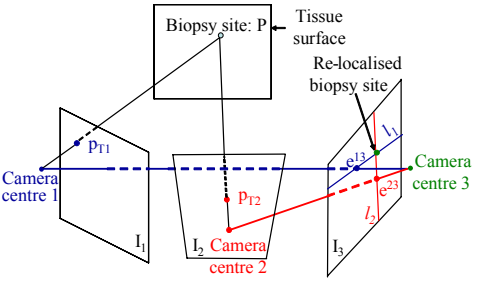


Fig. 1. Method for biopsy site re-localisation with three images and two epipolar lines

epipolar line $\boldsymbol{l_2}$ can be defined similarly for $\mathbf{F_{23}p_{T2}}$. As $\mathbf{p_{T1}}$ and $\mathbf{p_{T2}}$ correspond to the same biopsy site location in the 3D-space, the intersection of $\boldsymbol{l_1}$ and $\boldsymbol{l_2}$ returns the location of the biopsy site in image $\mathbf{I_3}$.

2.2 Re-localisation System

The two epipolar lines are determined with computations described in [8]. As summarised in Fig. 2, two pairs of endoscopic images are provided as inputs to the system and processed independently until step 5.

The system is described as follows for Input 1:

Feature Tracking and Matching

As described in [9] for bronchoscopy, a block matching in step 1 tracks blocks of *M*x*N* pixels through a series of subsequent images to match features $\{\mathbf{p}_{1i}, \mathbf{p}_{3i}\}$ from images $\mathbf{I_1}$ and $\mathbf{I_3}$. The similarity between two blocks, $\mathbf{Bm}_i$ and $\mathbf{Bn}_i$, in images m and n is the square of the cross-correlation of each block with pixel intensities $\mathbf{Bm}_i(k,l)$ and $\mathbf{Bn}_i(k,l)$, average intensities $\mathbf{Bm}_{i0}$ and $\mathbf{Bn}_{i0}$, and standard deviations σ_{Bmi} and σ_{Bni}:

$$CC^2(\mathbf{Bm}_i,\mathbf{Bn}_i)=\frac{\left[\sum_{k=1}^{M}\sum_{l=1}^{N}(\mathbf{Bm}_i(k,l)-\mathbf{Bm}_{i0})(\mathbf{Bn}_i(k,l)-\mathbf{Bn}_{i0})\right]^2}{\sigma^2_{Bm_i}\sigma^2_{Bn_i}} \quad . \tag{2}$$

Inliers' Detection

The Maximum A Posteriori Sample Consensus (MAPSAC) has been presented as a robust method for inliers' detection among the features [10]. The error for matched features must be minimised, which is equivalent to minimising a cost function *C*:

$$C=\sum_{i=1..n}\rho\left(e_i^2\right) \text{ with } \rho\left(e_i^2\right)=\begin{cases} e_i^2 & \text{if } e_i^2<T \\ T & \text{if } e_i^2\geq T \end{cases} \quad . \tag{3}$$

As explained in [10], T is a threshold for the detection of inliers and e_i is the geometric distance for matched features $\{\mathbf{p}_{1i}, \mathbf{p}_{3i}\}$. The contribution of the inliers to the error and to the fundamental matrix computation is, therefore, taken into account.

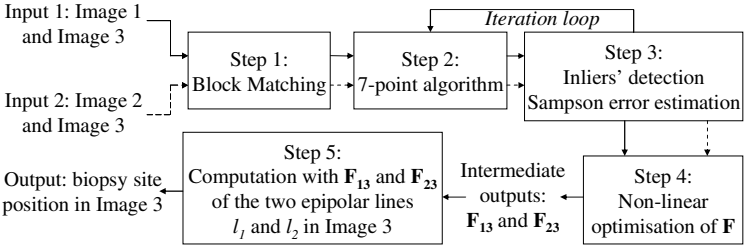


Fig. 2. Block diagram of the biopsy site re-localisation algorithm

Determination of the Epipolar Lines Intersection
Steps 2 and 3 are iteratively run over samples of 7 matched features. For each sample, the 7-point algorithm is applied [8] and returns one or three solutions for the fundamental matrix $\mathbf{F_{13}}$. The cost function C is applied with the Sampson distance e_i:

$$e_i^2 = \frac{\left(\mathbf{p}_{3i}^{T}\mathbf{F}_{13}\mathbf{p}_{1i}\right)^2}{\left(\mathbf{F}_{13}\mathbf{p}_{1i}\right)_1^2 + \left(\mathbf{F}_{13}\mathbf{p}_{1i}\right)_2^2 + \left(\mathbf{F}_{13}^{T}\mathbf{p}_{3i}\right)_1^2 + \left(\mathbf{F}_{13}^{T}\mathbf{p}_{3i}\right)_2^2} \cdot \quad (4)$$

$(\mathbf{F_{13}p}_{1i})_1$ is the 1st component of the vector $\mathbf{F_{13}p}_{1i}$. This error expresses how well $\mathbf{F_{13}}$ fits the matched features $\{\mathbf{p}_{1i}, \mathbf{p}_{3i}\}$. After the iteration loop, $\mathbf{F_{13}}$ minimizes C. A second estimation of $\mathbf{F_{13}}$, minimising the Sampson distance, is found in step 4 with a constrained non-linear optimisation applied to the inliers detected from the MAPSAC. Thus, in step 5, the required epipolar lines l_1 and l_2 for the biopsy site re-localisation are computed from $\mathbf{F_{13}}$ and $\mathbf{F_{23}}$ and return the expected result.

3 Validation

3.1 Materials and Method

Purpose. The validation consisted of first checking the recovery of the epipolar geometry to make sure that reliable epipolar lines could be computed and secondly of assessing the re-localisation error.

Data and Procedure. *In vivo* data were acquired with monocular endoscopes during gastroscopy of two patients. The same endoscopist introduced a 2 mm diameter miniprobe into the working channel of the endoscope, scanned the tissue surface, moved the endoscope camera to change the viewpoints, and removed the miniprobe. The re-localisation method was validated with 9 groups of 3 images from the 2 patients. Cardiac and breathing motion could be approximated by an affine transformation of the oesophagus surface.

Validation of the Recovery of the Epipolar Geometry. Endoscope motion usually comprises small rotations and translations around and along the optical axis of the endoscope in consecutive images. Thus, every time the re-localisation algorithm was applied, the computed epipoles were visually checked to ensure that they were in an

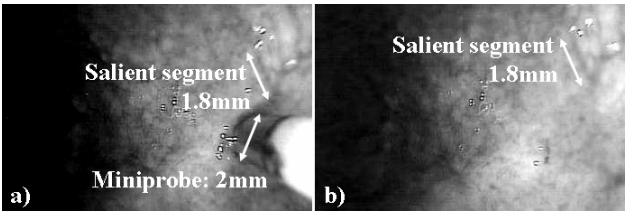


Fig. 3. Scales in the endoscopic images: a) The miniprobe helps compute the size of the vessel, b) which was used as a scale in Image 3

area in reasonable agreement with the endoscope motions. The total Sampson distance, defined as:

$$d = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} \frac{\left(\mathbf{p}_{3i}^{\mathrm{T}} \mathbf{F}_{13} \mathbf{p}_{1i}\right)^2}{\left(\mathbf{F}_{13}\mathbf{p}_{1i}\right)_1^2 + \left(\mathbf{F}_{13}\mathbf{p}_{1i}\right)_2^2 + \left(\mathbf{F}_{13}^{\mathrm{T}}\mathbf{p}_{3i}\right)_1^2 + \left(\mathbf{F}_{13}^{\mathrm{T}}\mathbf{p}_{3i}\right)_2^2}} \quad . \tag{5}$$

was computed for the fundamental matrices $\mathbf{F_{13}}$ and $\mathbf{F_{23}}$, obtained with: (1) the Normalised 8-Point (N8P) algorithm [8] applied to all the feature matches $\{\mathbf{p}_{1i}, \mathbf{p}_{3i}\}$ or $\{\mathbf{p}_{2i}, \mathbf{p}_{3i}\}$, and (2) with the re-localisation algorithm.

Estimation of the Re-localisation Error. This was carried out by the endoscopist. A biopsy site (a point on a salient vessel at the tissue surface or the miniprobe tip) was manually selected in Images 1 and 2. It was re-localised both visually in Image 3 to return a ground-truth location and automatically with the re-localisation algorithm. An error in pixels and in millimetres was determined from these two locations. As the diameter of the miniprobe was 2 mm, the size of a salient segment, e.g. a vessel, in the same plane as the miniprobe could be computed (Fig. 3 a)). This salient segment defined a scale in Image 3 for the conversion of the error from pixels to millimetres and for an estimation of the tissue dimensions in the Field Of View (FOV) of the image (Fig. 3 b)).

3.2 Results

Recovery of the Epipolar Geometry. As indicated in Fig. 4, the epipoles $\mathbf{e^{13}}$ and $\mathbf{e^{23}}$ computed for the pairs of images 1 and 3 (I 1-3), and 2 and 3 (I 2-3) were in an area in agreement with the endoscope motions. The re-localisation algorithm gave an order of magnitude reduction in the Sampson distance in comparison to the N8p algorithm. Thus, reliable epipolar geometry was recovered with the re-localisation algorithm.

Re-localisation: The estimated errors in pixels and in millimetres for the nine groups of three endoscopic images are presented in Table 1. This error varied from 2 pixels to 50 pixels in the x and y directions of the image. The endoscopist assessed the error for each group of images in millimetres with the scale that was provided in Image 3. In most cases, this error was smaller than 1.5 mm which corresponds to approximately 100-150 cells. In practice, an extracted tissue sample has a diameter of 5 mm; therefore, the endoscopist would have a high chance of extracting a part of the region that he previously analysed with the miniprobe using forceps. The errors greater than 1.5 mm

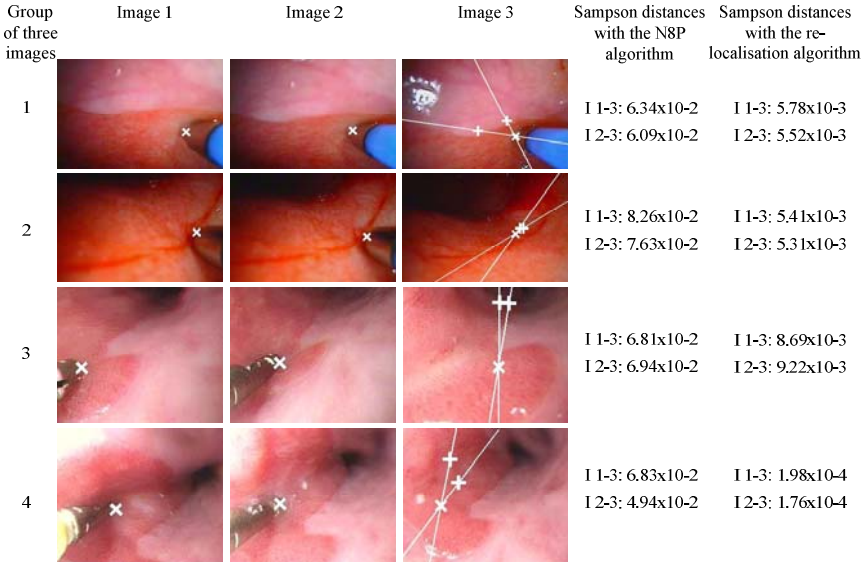


Fig. 4. Results for the recovery of the epipolar geometry and for the re-localisation of a biopsy site: the computed epipoles $\mathbf{e}^{13}$ and $\mathbf{e}^{23}$ are described with the sign '+' in Image 3 for each group of images. The sign 'x' describes the biopsy site, whose positions $\mathbf{p}_{T1}$ and $\mathbf{p}_{T2}$ are manually selected in Image 1 and Image 2. In Image 3, the sign 'x' describes the biopsy site computed with the two epipolar lines shown in Image 3. Two Sampson distances are computed in pixels for each group of images: one for the pair of images I 1-3, and the other for the pair I 2-3.

Table 1. Re-localisation errors in pixels and estimation of the absolute errors in millimeters

Image	Dimension (pixels)	FOV (mm)	Error (pixels)	Error (mm)
1	229 x 344	20 x 15	1.3 x 7.6	0.60
2	229 x 344	30 x 30	12.8 x 24.7	1.39
3	229 x 344	30 x 30	1.9 x 0.5	0.52
4	229 x 344	30 x 30	14.9 x 26	0.59
5	229 x 344	15 x 10	47.9 x 12.4	1.28
6	193 x 235	15 x 20	29.2 x 3.4	1.40
7	193 x 235	30 x 30	12.9 x 3.6	1.22
8	193 x 235	30 x 30	2.2 x 32.5	2.56
9	193 x 235	20 x 30	8.3 x 31.2	2.13

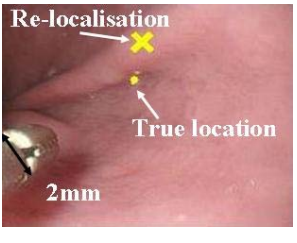


Fig. 5. Re-localisation error

become much more critical. They usually occur in endoscopic images with a poor contrast. In Fig. 5 for example, the tissue texture was smooth: no vessel or other salient structure existed in this region. Therefore, the main features extracted with the block matching were localised in the tissue folds, or in zones with a weak contrast. The features were sensitive to the endoscope motions and to the changes in illumination. Several of them were inaccurately tracked by a few pixels but were not detected as outliers since they did not create large errors. Thus, for these cases, the recovery of

the epipolar geometry was inaccurate and the epipolar lines derived from the biopsy site in Image 1 and 2 did not pass through the actual location of the biopsy site in Image 3.

4 Conclusion

A method for the re-localisation of biopsy sites in endoscopic images has been described. It has the advantages of using epipolar geometry properties without directly computing the rotations and the translations of the endoscope camera. This method needs only the determination of the fundamental matrix to compute the necessary epipolar lines. Clinical data validation showed re-localisation can be accurate to less than 1.5 mm or to within 100-150 cells as determined by the direct scaling of images and independent corroboration by an endoscopist. Although tested only with gastroscopic images, this method should work equally well for other endoscopic applications. Regions with greater structure and contrast would improve the ease and accuracy of the final re-localisation results. The technique is robust as satisfactory results were obtained even with substantial camera motions between video images.

Future work includes refining the re-localisation error by improving (a) both the feature tracking and matching, and (b) outlier detection. The re-localisation algorithm may have greater difficulty in cases where the endoscope returns to the biopsy site region at a later time during the examination or subsequent investigations. Thus, future work will also focus on (c) an extension of the method to the cases when a biopsy site needs to be re-localised with independent video sequences and not only successive video images from a single sequence. Finally, (d) a phantom study is required to estimate the re-localisation errors more accurately.

Acknowledgments. This work was undertaken at UCLH/UCL with funding from the Department of Health (UK) DLH Clinician Scientist Award to Dr. Cook, Engineering and Physical Sciences Research Council (EPSRC – grant reference EP/C523016/1), Department of Health's NIHR Comprehensive Biomedical Research Centre (UCLH), Cancer Research UK: grant to the UCL Experimental Cancer Medicine Centre. The views expressed in this publication are those of authors and not necessarily those of the funders.

References

1. Helferty, J.P., Sherbondy, A.J., Kiraly, A.P., et al.: Computer-based system for the virtual-endoscopic guidance of bronchoscopy. Computer Vision and Image Understanding 108(1-2), 171–187 (2007)
2. Mori, K., Deguchi, D., Ishitani, K., Kitasaka, T., Suenaga, Y., Hasegawa, Y., Imaizumi, K., Takabatake, H.: Bronchoscope tracking without fiducial markers using ultra-tiny electromagnetic tracking system and its evaluation in different environments. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 644–651. Springer, Heidelberg (2007)

3. Lovat, L.B., Johnson, K., Mackenzie, G.D., et al.: Elastic scattering spectroscopy accurately detects high grade dysplasia and cancer in Barrett's oesophagus. Gut. 55, 1078–1083 (2006)
4. Pohl, H., et al.: Miniprobe confocal laser microscopy for the detection of invisible neoplasia in patients with Barrett's oesophagus. Gut 57, 1648–1653 (2008)
5. Speidel, S., Delles, M., Gutt, C., et al.: Tracking of Instruments in Minimally Invasive Surgery for Surgical Skill Analysis. In: Yang, G.-Z., Jiang, T.-Z., Shen, D., Gu, L., Yang, J. (eds.) MIAR 2006. LNCS, vol. 4091, pp. 148–155. Springer, Heidelberg (2006)
6. Voros, S., Long, J.A., Cinquin, P.: Automatic Localization of Laparoscopic Instruments for the Visual Servoing of an Endoscopic Camera Holder. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 535–542. Springer, Heidelberg (2006)
7. Hu, M., Penney, G., Edwards, P., et al.: 3D reconstruction of internal organ surfaces for minimal invasive surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 68–77. Springer, Heidelberg (2007)
8. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2004)
9. Mori, K., Deguchi, D., Sugiyama, J., et al.: Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. Medical Image Analysis 6(3), 321–336 (2002)
10. Torr, P.H.S., Davidson, C.: IMPSAC: A synthesis of importance sampling and random sample consensus. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(3), 354–364 (2003)

Probabilistic Region Matching in Narrow-Band Endoscopy for Targeted Optical Biopsy

Selen Atasoy[1,2], Ben Glocker[1], Stamatia Giannarou[2], Diana Mateus[1], Alexander Meining[3], Guang-Zhong Yang[2], and Nassir Navab[1]

[1] Computer Aided Medical Procedures (CAMP), Technische Universität München
{atasoy,glocker,mateus}@cs.tum.edu
[2] Visual Information Processing Group, Imperial College London
{catasoy,stamatia.giannarou,g.z.yang}@imperial.ac.uk
[3] Department of Gastroenterology, Technische Universität München
alexander.meining@lrz.tu-muenchen.de

Abstract. Recent advances in biophotonics have enabled *in-vivo, in-situ* histopathology for routine clinical applications. The non-invasive nature of these optical 'biopsy' techniques, however, entails the difficulty of identifying previously visited biopsy locations, particularly for surveillance examinations. This paper presents a novel region-matching approach for narrow-band endoscopy to facilitate retargeting the optical biopsy sites. The task of matching sparse affine covariant image regions is modelled in a Markov Random Field (MRF) framework. The proposed model incorporates appearance based region similarities as well as spatial correlations of neighbouring regions. In particular, a geometric constraint that is robust to deviations in relative positioning of the detected regions is introduced. In the proposed model, the appearance and geometric constraints are evaluated in the same space (photometry), allowing for their seamless integration into the MRF objective function. The performance of the method as compared to the existing state-of-the-art is evaluated with both *in-vivo* and simulation datasets with varying levels of visual complexities.

1 Introduction

Oesophageal Adenocarcinoma (OAC) is the most rapidly increasing cancer in Europe and the United States, which has a 5-year survival rate of only 10% [1]. Barrett's Oesophagus (BO), referring to the abnormal change of the oesophageal mucosa caused by gastro-oesophageal reflux (Fig. 1a-b), is the only recognized precursor to OAC. Therefore, for patients diagnosed with BO, periodic surveillance by gastrointestinal (GI) endoscopy together with systematic biopsy is important for the early detection and prevention of OAC.

In current surveillance protocols, a new technique called Narrow-Band Endoscopic Imaging (NBI), has shown advantage compared to conventional white light endoscopy as it allows for detailed visualization of mucosa and the underlying vascular patterns (Fig. 1a-b). A further technique called Fibered Confocal Microscopy (FCM) which enables real-time visualization of cellular structures *in-vivo* and *in-situ* (Fig. 1c) is also introduced recently. During GI endoscopy, a fibered confocal microprobe

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 499–506, 2009.

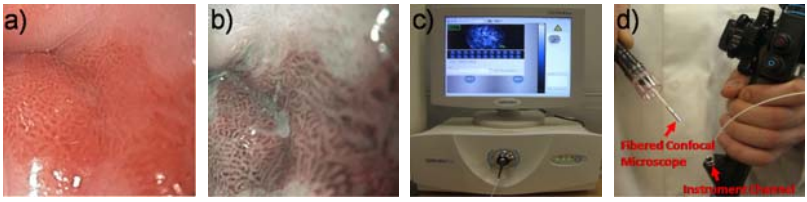


Fig. 1. Appearance of BO a) in white light endoscopy and b) in NBI. c) The FCM machine and d) FCM microprobe passing through the instrument channel of a standard endoscope.

can be inserted easily through the instrument channel of a standard endoscope (Fig. 1d), thus providing *in-situ* histopathology without the need for tissue biopsy [2]. This has significant benefits in terms of ease of examination, patient comfort and real-time feedback. In practice, however, the non-invasive nature of the procedure also makes it difficult to return to previously examined biopsy sites in surveillance endoscopy due to the absence of scar on the tissue. The purpose of this paper is to present a novel image-based region matching method for biopsy site re-targeting in NBI.

Region matching in NBI entails several challenges, which include tissue deformation, prevalence of similar surface textures and mucosal patterns. As the endoscope is very close to the tissue, small differences in the visible scales of the same feature can cause a significant change in the visual content. Furthermore, the common issue of view-invariant scene matching also needs to be addressed.

Viewpoint invariant scene matching is a well studied problem in computer vision and it typically proceeds by representing the scene as a collection of affine covariant regions which are described by a vector computed from the regions' appearances. Usually a nearest neighbour matching of the descriptor vectors incorporating geometric constraints is used to eliminate possible outliers ([3-5] and references within). In endoscopy, the major focus is directed towards short-baseline matching/tracking in the presence of tissue deformation [6].

Recently, spectral methods have been proposed for region-matching in images undergoing non-rigid transformations [7-9]. These methods model a graph for the feature set in each image and estimate their correspondences by graph matching. Thereby the geometric relations are modelled in terms of point locations where distance and/or orientation preservation is enforced. The main focus of these approaches lies on estimating the optimal solution for the NP-hard graph matching problem rather than on optimal modelling. However, Caetano *et al.* have demonstrated that finding the (near) optimal graphs can greatly simplify the matching problem and improve the results [10]. The authors proposed a learning based approach for optimal graph extraction. An MRF model [11] and a graph matching approach [8] with optimal model parameter learning are also presented for the correspondence problem.

In this paper, we will focus on deriving the (near) optimal MRF model for the feature correspondence problem in NBI. The proposed model incorporates appearance based region similarities as well as the spatial correlations of neighbouring regions. To this end, we introduce a geometric constraint that evaluates the consistency between neighbouring matches on their photometric properties. Evaluation of the appearance and geometric constraints in the same space (photometry) allows for their seamless integration. The performance of the proposed method is evaluated with both *in-vivo* and simulation datasets.

2 Methods

The proposed method involves initial affine covariant region detection. This is followed by formulating a particular MRF model for the matching problem. Finally, the optimal labelling is computed using Belief Propagation.

2.1 Region Detection and Description

Affine covariant regions are detected independently on both images using affine invariant anisotropic region detector [12], which is shown to be robust against small deformations. For viewpoint invariant description, each elliptical region p is normalized by the corresponding affine transformation M_p (determined by the shape of the ellipse) and mapped onto the corresponding circular region $\overline{\mathrm{p}} = \mathrm{M}_\mathrm{p} \cdot \mathrm{p}$ (Fig. 2a). Then, the dominant gradient orientation $\vartheta_{\overline{\mathrm{p}}}$ is estimated from the local image gradients and the SIFT descriptor [3] $d(\overline{\mathrm{p}}, \sigma_{\overline{\mathrm{p}}}, \vartheta_{\overline{\mathrm{p}}})$ is computed from the circular patch $\overline{\mathrm{p}}$ using the characteristic scale $\sigma_{\overline{\mathrm{p}}}$ and the dominant gradient orientation $\vartheta_{\overline{\mathrm{p}}}$.

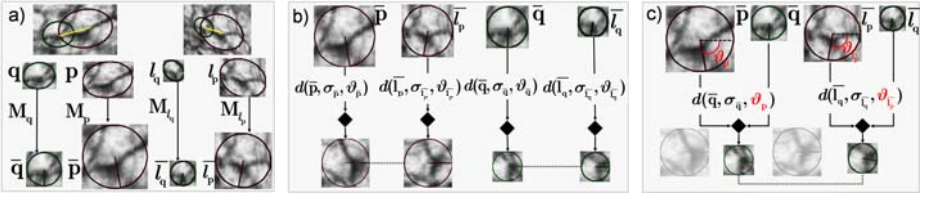


Fig. 2. a) Viewpoint invariant region description. b) Unary costs computed from the region descriptors, where the diamond indicates the SIFT descriptor computed on the affine normalized patches. c) The proposed pair-wise costs. SIFT descriptors are computed on the patches $\overline{\mathrm{q}}$ and $\overline{l}_\mathrm{q}$ using the dominant gradient orientations ϑ_p, ϑ_{l_p} of the regions $\overline{\mathrm{p}}$ and $\overline{l}_\mathrm{p}$. Two compared image patches are the same, whereas the length and orientation of the line segment between two region centres are not preserved as illustrated via the yellow lines in a).

2.2 Matching through Markov Random Fields

Given the computed region descriptors, we model the matching problem as global optimization of an MRF labelling. We define the regions in the first image to be the nodes $\mathcal{G} = \{1,..,\mathrm{n}\}$ of the MRF and the regions in the second image to be the labels $\mathcal{L}^+ = \{l_0, l_1, .., l_M\}$ including the null-label l_0, which is assigned to regions without true correspondence in the second image. In this paper, we consider only up to pair-wise relations. Thus, finding the maximum a posteriori (MAP) estimate of the optimum labelling $\mathbf{l}^*$ is equivalent to minimizing the energy function:

$$E_{\mathrm{MRF}}(\mathbf{l}) = \sum_{\mathrm{p}\in\mathcal{G}} V_\mathrm{p}(l_\mathrm{p}) + \sum_{\mathrm{p}\in\mathcal{G}} \sum_{\mathrm{q}\in\mathcal{N}(\mathrm{p})} V_{\mathrm{pq}}(l_\mathrm{p}, l_\mathrm{q}) \tag{1}$$

where $V_\mathrm{p}(l_\mathrm{p})$ is the unary cost of assigning the label l_p to the node p, $V_{\mathrm{pq}}(l_\mathrm{p}, l_\mathrm{q})$ is the pair-wise cost and $\mathcal{N}$ defines the neighbourhood system.

2.3 Unary Costs

In our model, photometric similarities between the node and the label regions are evaluated via the unary costs by defining $V_{\mathrm{p}}(l_{\mathrm{p}})$ to be the distance of the SIFT descriptors of the node $\overline{\mathrm{p}}$ and label $\overline{l_{\mathrm{p}}}$ regions (Fig. 1b). We further define the cost $V_{\mathrm{p}}(l_0)$ of assigning the null-label l_0 to a node p to be a function of the photometric similarities. The motivation is that assigning the null-label l_0 to a region that has a strong correspondence in the second image should have a higher cost than assigning it to a region with no (strong) correspondence. We define the *null-cost function* of the node p as $V_{\mathrm{p}}(l_0) = \alpha \cdot (1 - \min(V_{\mathrm{p}}(\cdot)))$, where $\min(V_{\mathrm{p}}(\cdot))$ is the minimum cost of assigning a label to the node p, and α is the factor regulating the trade-off between the quality and the number of matches. (For all our *in-vivo* datasets, the best performance is achieved for $\alpha = 0.5$). The final unary costs are computed as:

$$V_{\mathrm{p}}(l_{\mathrm{p}}) = \begin{cases} \arccos(d(\overline{\mathrm{p}}, \sigma_{\overline{\mathrm{p}}}, \vartheta_{\overline{\mathrm{p}}}) \cdot d(\overline{l_{\mathrm{p}}}, \sigma_{\overline{l_{\mathrm{p}}}}, \vartheta_{\overline{l_{\mathrm{p}}}})) / \arccos(0) & \text{if } l_{\mathrm{p}} \neq l_0 \\ \alpha \cdot (1 - \min(V_{\mathrm{p}}(\cdot))) & \text{otherwise} \end{cases}, \quad (2)$$

where all costs $V_{\mathrm{p}}(\cdot)$ are normalized to the interval $[0,1]$ by dividing by the maximum possible angle between two descriptor vectors; $\arccos(0)$.

2.4 Neighbourhood Systems

In the context of the matching problem, each region is allowed to have at most one correspondence in the second image, *i.e.*, each label can be assigned at most to one node. This uniqueness constraint is included into the energy function by connecting each node with all the other nodes within the *global neighbourhood system* $\mathcal{N}$ and by defining the pair-wise cost for assigning the same label to two nodes to be infinite: $V_{\mathrm{pq}}(l_{\mathrm{p}}, l_{\mathrm{q}}) = \infty$ if $l_{\mathrm{p}} = l_{\mathrm{q}} \neq l_0$. We further define a *local neighbourhood system* in order to impose flexible local geometric constraints. This neighbourhood system is defined for both the nodes and the labels to impose neighbourhood preservation as the initial geometric constraint (Eq. 3.1). The local neighbourhood $\mathcal{N}_{\mathrm{local}}(\mathrm{p})$ of a region p is set to be: $\mathcal{N}_{\mathrm{local}}(\mathrm{p}) = \{\mathrm{q} \neq \mathrm{p} \mid \; || \mathrm{p} - \mathrm{q} || < t\}$, where $|| \mathrm{p} - \mathrm{q} ||$ is the Euclidean distance between the centres of p and q and t is a threshold value. (We use $t = 10\%$ and $t = 20\%$ of the image size for the node- and label-neighbourhoods respectively to ensure the connectivity of two neighbouring regions after a large viewpoint change). For regions fulfilling the neighbourhood preservation, a novel geometric constraint is imposed measuring the consistency of two matches.

2.5 Pair-Wise Costs

In this paper, we propose a geometric constraint based on the assumption that neighbouring regions move with similar transformations. The idea is as follows: if two neighbouring regions p and q have the corresponding regions l_{p} and l_{q} in the second image, then there exist two affine transformations A_{p} and A_{q} such;

$l_{\mathrm{p}}(\mathbf{x}) = \mathrm{A}_{\mathrm{p}} \cdot \mathrm{p}(\mathbf{x}) = s_{\mathrm{p}} \cdot \mathrm{R}_{\mathrm{p}} \cdot \mathrm{M}_{\mathrm{p}} \cdot \mathrm{p}(\mathbf{x})$ and $l_{\mathrm{q}}(\mathbf{x}) = \mathrm{A}_{\mathrm{q}} \cdot \mathrm{q}(\mathbf{x}) = s_{\mathrm{q}} \cdot \mathrm{R}_{\mathrm{q}} \cdot \mathrm{M}_{\mathrm{q}} \cdot \mathrm{q}(\mathbf{x})$ where s_{p} and s_{q} are scale factors. Theoretically, for spatially close regions on the same plane it holds $\mathrm{A}_{\mathrm{p}} = \mathrm{A}_{\mathrm{q}}$. However, for neighbouring regions on different planes this assumption is too restrictive and can be relaxed by assuming only $\mathrm{R}_{\mathrm{p}} = \mathrm{R}_{\mathrm{q}} = \mathrm{R}$, where R is the rotation of the local neighbourhood between two images.

If two neighbouring matches $m_{\mathrm{p}} = (\mathrm{p}, l_{\mathrm{p}})$ and $m_{\mathrm{q}} = (\mathrm{q}, l_{\mathrm{q}})$ are true correspondences, then the SIFT descriptors $d(\overline{\mathrm{q}}, \sigma_{\overline{\mathrm{q}}}, \vartheta_{\overline{\mathrm{p}}})$ and $d(\overline{l_{\mathrm{q}}}, \sigma_{\overline{l_{\mathrm{q}}}}, \vartheta_{\overline{l_{\mathrm{p}}}})$ computed on the patches $\overline{\mathrm{q}}$ and $\overline{l_{\mathrm{q}}}$ using their own characteristic scales $\sigma_{\overline{\mathrm{q}}}$, $\sigma_{\overline{l_{\mathrm{q}}}}$ and the dominant gradient orientations $\vartheta_{\overline{\mathrm{p}}}$, $\vartheta_{\overline{l_{\mathrm{p}}}}$ of $\overline{\mathrm{p}}$ and $\overline{l_{\mathrm{p}}}$ should be similar as the local rotation can be determined as $\mathrm{R} = \vartheta_{\overline{l_{\mathrm{p}}}} - \vartheta_{\overline{\mathrm{p}}}$ (Fig. 1c). (Recall that $\mathrm{M}_{\mathrm{p}} \cdot \mathrm{p} = \overline{\mathrm{p}}$ and $\mathrm{M}_{\mathrm{q}} \cdot \mathrm{q} = \overline{\mathrm{q}}$). This similarity measure indicates the consistency of the matches m_{p} and m_{q}, as the rotation estimated from the match m_{p} is evaluated on the regions of the match m_{q}. Combining with the neighbourhood preservation, the pair-wise costs to evaluate geometric constraints are defined as:

$$\psi_{\mathrm{pq}}(l_{\mathrm{p}}, l_{\mathrm{q}}) = \begin{cases} \infty & \text{if } (l_{\mathrm{p}} \notin \mathcal{N}_{\mathrm{local}}(l_{\mathrm{q}})) \quad (3.1) \\ \arccos(d(\overline{\mathrm{q}}, \sigma_{\overline{\mathrm{q}}}, \vartheta_{\overline{\mathrm{p}}}) \cdot d(\overline{l_{\mathrm{q}}}, \sigma_{\overline{l_{\mathrm{q}}}}, \vartheta_{\overline{l_{\mathrm{p}}}})) / \arccos(0) & \text{if } (l_{\mathrm{p}} \in \mathcal{N}_{\mathrm{local}}(l_{\mathrm{q}})) \quad (3.2) \end{cases}.$$

Introducing the geometric constraints the final pair-wise costs are defined as:

$$V_{\mathrm{pq}}(l_{\mathrm{p}}, l_{\mathrm{q}}) = \begin{cases} \infty & \text{if } (l_{\mathrm{p}} = l_{\mathrm{q}} \neq l_0) \\ \alpha & \text{if } (l_{\mathrm{p}} = l_0) \vee (l_{\mathrm{q}} = l_0) \\ \psi_{\mathrm{pq}}(l_{\mathrm{p}}, l_{\mathrm{q}}) & \text{if } (\mathrm{q} \in \mathcal{N}_{\mathrm{local}}(\mathrm{p})) \\ 0 & \text{otherwise} \end{cases}, \quad (4)$$

where α is the same factor used in the unary null-costs.

The derived geometric constraint (Eq. 3.2) is invariant to changes in scale and relies only on the assumption of similar rotations within a local neighbourhood. This is in contrast to previous methods taking global consistency of the features into account. Such methods would fail in case of global deformation, which is present in our applications. This constraint also allows us to be locally more flexible than recent methods making stronger assumptions, such as invariance of the distance between the neighbouring points or the orientation of line segment connecting their centroids [7-9]. Furthermore, the proposed constraint evaluates local image geometry based on the photometric properties of the patches rather than their spatial locations. This allows for the evaluation of the unary and the pair-wise costs in the same space and combining them in the objective function without using any weighting parameters.

2.6 MAP Estimation

The MAP labelling of the proposed MRF model is estimated using Belief Propagation (BP) [13]. The non-submodularity of the pair-wise costs restricts the choice of the

MRF inference algorithms to those without prior constraints on the class of energy functions. In this paper, without loss of generality we use the BP algorithm.

3 Experiments and Results

The performance of the proposed method is evaluated on 4 *in-vivo* and 4 simulation datasets and compared to 3 matching strategies evaluated in [14]. The regions are detected and described as explained in Section 3.1. The threshold-based (TB) [14], nearest-neighbour (NN) [14] and the nearest neighbour distance ratio matching (NNDR) [3,14] are applied for varying threshold values. MRF-based method is performed for different values of the factor α within the convergence range of the optimization. We further compared the hypergraph matching algorithm (HGM) using the proposed affine invariant geometric measure via quadripartite point relations [7]. However, these graph matching methods are not adapted to large number of non-matching regions (43%-87% in our datasets). Therefore, the performance of the HGM is poor and not illustrated here in detail. For quantitative analysis, we evaluate $recall$ (the ratio of correct matches to the total number of correspondences) versus $1-precision$ (the number of false matches with respect to the number of matched regions). For the best matching results $recall = 1$ and $1-precision = 0$.

3.1 Simulation Studies

For evaluation with known ground truth data, we created 4 simulation datasets (2 for viewpoint change and 2 for deformation). In the first study, we generated images under different viewpoint conditions by transforming 2 *in-vivo* images (one veined and one structured area) with known homographies. In the second study, we deformed 2 *in-vivo* images (one structured and one homogenous area) and tracked the detected regions. In both studies, two regions were accepted as a correct match if the distance between the centres of the transformed and detected ellipses was less than 1% of the image size and the overlap was more than 55%. Figs. 4(a-c) demonstrate that MRF matching results in a better performance than all compared methods for structured scenes. Fig. 4d shows that in the presence of non-distinctive regions, MRF-matching and NNDR (which favours distinctive matches) exhibit a similar performance. The matching results of the MRF model are presented in Figs. 3(a-d).

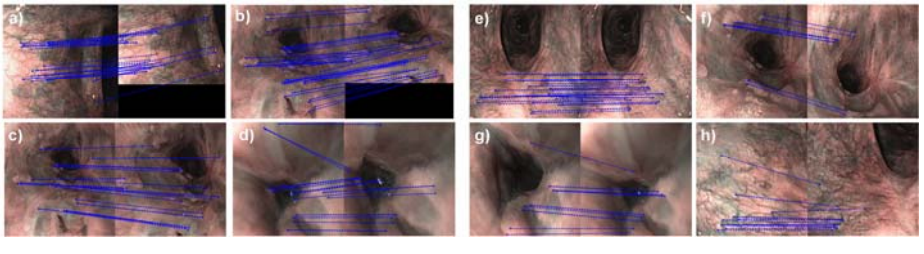


Fig. 3. Matching results of the MRF model on a-b) viewpoint change c-d) deformation simulation datasets, as well as e) the first, f) second, g) third and h) fourth *in-vivo* datasets

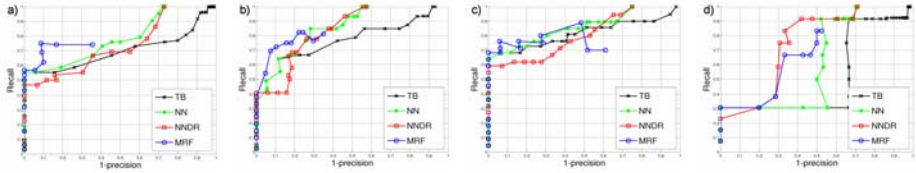


Fig. 4. Validation of the results on simulation datasets. Viewpoint change on images of a) veined b) structured tissue. Deformation on images of c) structured and d) homogenous tissue.

3.2 In-vivo Studies

For the *in-vivo* studies, we used 4 NBI datasets with different viewpoint and photometric conditions. The first 3 datasets contain two distant frames of the same GI procedure from different viewpoints showing a veined area (Fig. 3e), structured area (Fig. 3f) and homogenous area with large deformation (Fig. 3g). The fourth dataset contains images acquired during two different GI examinations with a time difference of 3 months where the patient underwent chemotherapy. This results in large changes in the visual appearance of the tissue (Fig. 3h). For the *in-vivo* data sets, the ground truth data was provided by manual labelling. Fig. 5 demonstrates that for all *in-vivo* cases the proposed MRF model performs better than the state-of-the art descriptor matching techniques. For all datasets (simulation and *in-vivo*), maximum recall values for the acceptable precision interval (80%-100% inliers) are summarized in Table1. The matching results for the *in-vivo* datasets are presented in Fig 3e-f.

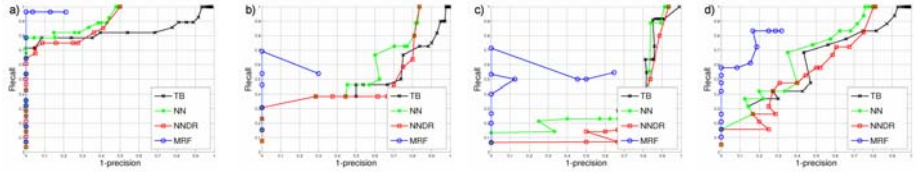


Fig. 5. Validation of the results on *in-vivo* datasets. a-b-c-d) show the recall versus precision of each matching algorithm for the first, second, third and fourth in-vivo dataset, respectively.

Table 1. Summary of the maximum recall values for the precision interval [0.8-1.0] (80%-100% inliers) for the simulation and *in-vivo* datasets

	Viewp.1	*Viewp.2*	*Def.1*	*Def.2*	*In-vivo1*	*In-vivo2*	*In-vivo3*	*In-vivo4*
MRF	0.75	0.75	0.76	0.31	0.96	0.69	0.71	0.83
NN	0.58	0.66	0.73	0.31	0.82	0.31	0.13	0.37
NNDR	0.53	0.68	0.62	0.31	0.75	0.31	0.06	0.21
TB	0.55	0.63	0.73	0.31	0.78	0.31	0.13	0.32

4 Conclusion

In this paper, we have investigated the task of region matching for NBI and propose a new method towards an image-based solution for consistent re-targeting of optical

biopsy sites. To this end, we present an MRF model for matching affine covariant regions incorporating a novel geometric constraint for dealing with large changes in the observed datasets. Our results demonstrate the robustness of the proposed model for deformable wide-baseline matching on *in-vivo* and simulation datasets. For future work, we plan to further extend our approach in order to provide a complete framework for this novel and challenging application. This would require both detection and tracking taking into account the sequential appearances of features within the first and secondary examinations.

Acknowledgments. The authors would like to thank Stefan Hinterstoisser, Douglas Mcilwraith and Peter Mountney for constructive discussions.

References

1. Wani, S., Sharma, P.: The rationale for screening and surveillance of Barrett's metaplasia. Best Practice & Research Clinical Gastroenterology 20, 829–842 (2006)
2. Meining, A., Saur, D., Bajbouj, M., Becker, V., Peltier, E., Höfler, H., von Weyhern, C., Schmid, R., Prinz, C.: In Vivo Histopathology for Detection of Gastrointestinal Neoplasia With a Portable. Confocal Miniprobe: An Examiner Blinded Analysis Clinical Gastroenterology and Hepatology 5, 1261–1267 (2007)
3. Lowe, D.: Distinctive Image Features from Scale-Invariant Keypoints. Int. Journal of Computer Vision, 91–110 (2004)
4. Sivic, J., Zisserman, A.: Video Google: a text retrieval approach to object matching in videos. In: ICCV, pp. 1470–1477 (2003)
5. Schaffalitzky, F., Zisserman, A.: Automated location matching in movies. Computer Vision and Image Understanding 92, 236–264 (2003)
6. Mountney, P., Lo, B., Thiemjarus, S., Stoyanov, D., Zhong-Yang, G.: A Probabilistic Framework for Tracking Deformable Soft Tissue in Minimally Invasive Surgery. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 34–41. Springer, Heidelberg (2007)
7. Zass, R., Shashua, A.: Probabilistic graph and hypergraph matching. In: CVPR (2008)
8. Torresani, L., Kolmogorov, V., Rother, C.: Feature correspondence via graph matching: Models and global optimization. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 596–609. Springer, Heidelberg (2008)
9. Leordeanu, M., Hebert, M.: A spectral technique for correspondence problems using pairwise constraints. In: ICCV, pp. 1482–1489 (2005)
10. Caetano, T., Cheng, L., Le, Q., Smola, A.: Learning graph matching. In: ICCV (2007)
11. Li, S.Z.: A Markov random field model for object matching under contextual constraints. In: CVPR, pp. 866–869 (1994)
12. Giannarou, S., Visentini-Scarzanella, M., Yang, G.Z.: Affine-Invariant Anisotropic Detector For Soft Tissue Tracking in Minimally Invasive Surgery. In: ISBI (2009)
13. Pearl, J.: Probabilistic Reasoning. Morgan Kaufmann, San Francisco (1988)
14. Mikolajczyk, K., Schmid, C.: A Performance Evaluation of Local Descriptors. In: PAMI, pp. 1615–1630 (2005)

Tracked Regularized Ultrasound Elastography for Targeting Breast Radiotherapy

Hassan Rivaz, Pezhman Foroughi, Ioana Fleming, Richard Zellars, Emad Boctor, and Gregory Hager

Johns Hopkins University, Baltimore, MD, USA

Abstract. Tracked ultrasound elastography can be used for guidance in partial breast radiotherapy by visualizing the hard scar tissue around the lumpectomy cavity. For clinical success, the elastography method needs to be robust to the sources of decorrelation between ultrasound images, specifically fluid motions inside the cavity, change of the appearance of speckles caused by compression or physiologic motions, and out-of-plane motion of the probe. In this paper, we present a novel elastography technique that is based on analytic minimization of a regularized cost function. The cost function incorporates similarity of RF data intensity and displacement continuity, making the method robust to small decorrelations present throughout the image. We also exploit techniques from robust statistics to make the method resistant to large decorrelations caused by sources such as fluid motion. The analytic displacement estimation works in real-time. Moreover, the tracked data, used for targeting the radiotherapy, is exploited for discarding frames with excessive out-of-plane motion. Simulation, phantom and patient results are presented.

1 Introduction

Breast irradiation after lumpectomy significantly reduces the risk of cancer recurrence. There is growing evidence suggesting that irradiation of only the involved area of the breast, partial breast irradiation (PBI), is as effective as whole breast irradiation [1]. Benefits of PBI include significantly shortened treatment time and fewer side effects as less tissue is treated. However, these benefits cannot be realized without localization of the lumpectomy cavity. Tracked ultrasound elastography can be used for localizing the lumpectomy cavity in the treatment room, minimizing tissue motion from planning to treatment.

This paper is focused on freehand palpation elastography, which involves estimating the displacement field of the tissue undergoing slow compression. Most elastography techniques estimate the displacement field using local cross correlation analysis of echoes [2,3,4]. These methods are very sensitive and accurate for calculating small displacements. However, elastography is subject to speckle decorrelation caused by various sources such as motion of subresolution scatterers, out-of-plane motion, high compression and complex fluid motions.

The prior of tissue deformation continuity can be used to make elastography more robust to signal decorrelation. Previous work on regularized elastography

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 507–515, 2009.

is computationally expensive [5,6]. Dynamic programming (DP) can be used to speed the optimization procedure [7], but it only gives integer displacements. Subpixel displacement estimation is possible [7], but it is computationally expensive if a fine subpixel level is desired. In addition, a fixed regularization weight is applied throughout the image. However, while two ultrasound images may correlate well in most parts, they can have small correlation in specific parts. Four examples of low correlation are: (1) correlation decreases with depth mainly due to a decrease in the ultrasonic signal to noise ratio, (2) correlation is low close to arteries due to complex motion and inside vessels due to blood motion, (3) correlation is extremely low in lesions that contain liquid due to the incoherent fluid motion [8,3], and (4) out-of-plane motion of movable structures within the image [8] causes low local correlation. To prevent such regions from introducing errors in the displacement estimation one should use large weights for the regularization term, resulting in over-smoothing.

Freehand palpation elastography provides ease-of-use and requires minimum additional cost. However, out-of-plane motion cannot be avoided in freehand palpation, which reduces the quality of any elastography method. Assisted freehand elastography [9] significantly reduces the out-of-plane motion but it requires addition of a device to the probe. Quality metrics such as persistence in strain images have also been developed to address this problem [10]. To measure the persistence, elastography is performed on two pairs of images and the resulting strain images are correlated. This method requires strain images for calculating the quality metric. Therefore, trying all the combinations in a series of frames to find the best pair for elastography will be computationally expensive.

In this paper, we present a novel elastography method based on analytic minimization (AM) of a cost function that incorporates similarity of echo amplitudes and displacement continuity. We introduce a novel regularization term and demonstrate that it minimizes displacement underestimation caused by smoothness constraint. We also introduce the use of robust statistics implemented via iterated reweighted least squares (IRLS) to treat uncorrelated ultrasound data as outliers. And finally, we use the tracking information to select the best pairs of frames for elastography. Simulation, phantom and patient experiments are presented for validation.

2 Regularized Displacement Estimation

Dynamic Programming (DP). DP is a discrete efficient optimization technique for causal systems. In DP elastography [7], a cost function is defined as

$$C(i,d_i) = \min_{d_{i-1}} \{C(i-1,d_{i-1}) + \alpha_a R(d_i,d_{i-1})\} + |I_1(i) - I_2(i+d_i)|\,,\ i = 2\cdots m \tag{1}$$

where d_i is the displacement of sample i, $R(d_i,d_{i-1}) = (d_i - d_{i-1})^2$ is an axial regularization term (axial, lateral and out-of-plane directions are respectively z, x and y in Figure 2 (a)), α_a is a weight for the regularization, I_1 and I_2 are corresponding RF-lines of before and after deformation and m is the length

of RF-lines. The cost function is minimized at $i = m$ and the d_i values that have minimized the cost function are traced back to $i = 1$, giving the d_i for all samples. We have implemented a 2D DP algorithm similar to [7] to generate integer displacements as a starting point for the next step of our algorithm.

Analytic Minimization (AM). We now propose a method that analytically minimizes a regularized cost function and gives the refined displacement field. Only axial displacements will be refined for strain calculation.

Having the integer displacements d_i from DP, it is desired to find Δd_i values such that $d_i + \Delta d_i$ gives the value of the displacement at the sample i for $i = 1 \cdots m$. Such Δd_i values will minimize the following regularized cost function

$$C\ (\Delta d_1, \cdots, \Delta d_m) = \Sigma_{i=1}^{m}\ [I_1(i) - I_2(i + d_i + \Delta d_i)]^2 + \alpha_a(d_i + \Delta d_i - d_{i-1} - \Delta d_{i-1})^2 + \alpha_l(d_i + \Delta d_i - d_i^{p\cdot} - \Delta d_i^{p\cdot})^2 \quad (2)$$

where superscript $p\cdot$ refers to the previous RF-line (adjacent RF-line in the lateral direction) and α_l is a weight for lateral regularization. Substituting $I_2(i + d_i + \Delta d_i)$ with its first order Taylor expansion approximation around d_i, we have

$$C\ (\Delta d_1, \cdots, \Delta d_m) = \Sigma_{i=1}^{m}\ [I_1(i) - I_2(i + d_i) - I_2'(i + d_i)\Delta d_i)]^2 + \alpha_a(d_i + \Delta d_i - d_{i-1} - \Delta d_{i-1})^2 + \alpha_l(d_i + \Delta d_i - d_i^{p\cdot} - \Delta d_i^{p\cdot})^2 \quad (3)$$

where I_2' is the derivative of the I_2. The optimal Δd_i values occur when the partial derivative of C w.r.t. Δd_i is zero. Setting $\frac{\partial C}{\partial \Delta d_i} = 0$ we have

$$({\mathbf{I}_2'}^2 + \alpha_a\mathbf{D} + \alpha_l\hat{\mathbf{I}})\Delta\mathbf{d} = \mathbf{I}_2'\mathbf{e} - (\alpha_a\mathbf{D} + \alpha_l\hat{\mathbf{I}})\mathbf{d} + \alpha_l\mathbf{d}^{t.p.\cdot},\quad \mathbf{D} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \cdots & 0 \\ \vdots & & & \ddots & \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix} \quad (4)$$

where $\mathbf{I}_2' = diag(I_2'(1+d_1) \cdots I_2'(m+d_m))$, $\Delta\mathbf{d} = [\Delta d_1 \cdots \Delta d_m]^T$, $\mathbf{e} = [e_1 \cdots e_m]^T$, $e_i = I_1(i) - I_2(i + d_i)$, $\mathbf{d} = [d_1 \cdots d_m]^T$, $\mathbf{d}^{t.p.\cdot} = \mathbf{d}^{p\cdot} + \Delta\mathbf{d}^{p\cdot}$ is the vector of total displacement of the previous line and $\hat{\mathbf{I}}$ is the identity matrix. $\mathbf{I}_2'$, $\mathbf{D}$ and $\hat{\mathbf{I}}$ are matrices of size $m \times m$ and $\Delta\mathbf{d}$, $\mathbf{r}$, $\mathbf{d}$ and $\mathbf{d}^{t.p.\cdot}$ are vectors of size m.

Biasing the Regularization. The regularization term $\alpha_a(d_i + \Delta d_i - d_{i-1} - \Delta d_{i-1})^2$ penalizes the difference between $d_i + \Delta d_i$ and $d_{i-1} + \Delta d_{i-1}$, and therefore can result in underestimation of the displacement field. Such underestimation can be prevented by biasing the regularization by ϵ to $\alpha_a(d_i + \Delta d_i - d_{i-1} - \Delta d_{i-1} - \epsilon)^2$, where $\epsilon = (d_m - d_1)/(m-1)$ is the average displacement difference between samples i and $i-1$. An accurate enough estimate of $d_m - d_1$ is known from the previous line. With the bias term, the R.H.S. of Equation 4 becomes $\mathbf{I}_2'\mathbf{e} - (\alpha_a\mathbf{D} + \alpha_l\hat{\mathbf{I}})\mathbf{d} + \alpha_l\mathbf{d}^{t.p.\cdot} + \mathbf{b}$ where the bias term is $\mathbf{b} = \alpha_a[-\epsilon\ \ 0 \cdots 0\ \ \epsilon]^T$ and all other terms are as before. Interestingly, except for the first and the last equation in this system, all other $m-2$ equations are same as Equation 4.

Equation 4 can be solved for $\Delta\mathbf{d}$ in $4m$ operations since the coefficient matrix ${\mathbf{I}_2'}^2 + \alpha_a\mathbf{D} + \alpha_l\hat{\mathbf{I}}$ is tridiagonal. Utilizing its symmetry, the number of operations

can be reduced to $2m$. The number of operations required for solving a system with a full coefficient matrix is more than $m^3/3$, significantly more than $2m$.

Making Tracking Resistant to Outliers. Even with pure axial compression, some regions of the image may move out of the imaging plane and increase the decorrelation. In such parts the confidence of the data term is less and therefore the weight of the regularization term should be increased. The parts of the image with low correlation can be regarded as outliers and therefore a robust estimation technique can limit their effect. Before deriving a robust estimator for $\Delta\mathbf{d}$, we rewrite Equation 3 as $C(\Delta\mathbf{d}) = \Sigma_{i=1}^{m}\rho(r_i) + R(\Delta\mathbf{d})$ where $r_i = I_1(i) - I_2(i+d_i) - I_2'(i+d_i)\Delta d_i$, $\rho(r_i) = r_i^2$ and R is the regularization term. The M-estimate of $\Delta\mathbf{d}$ is $\Delta\hat{\mathbf{d}} = \arg\min_{\Delta\mathbf{d}}\{\Sigma_{i=1}^{m}\rho(r_i) + R(\Delta\mathbf{d})\}$ where $\rho(u)$ is a robust loss function [11]. The minimization is solved by setting $\frac{\partial C}{\partial \Delta d_i} = 0$:

$$\rho'(r_i)\frac{\partial r}{\partial \Delta d_i} + \frac{\partial R(\Delta\mathbf{d})}{\partial \Delta d_i} = 0 \tag{5}$$

A common next step [11] is to introduce a weight function w, where $w(r_i).r_i = \rho'(r_i)$. This leads to a process known as "iteratively reweighted least squares" (IRLS), which alternates steps of calculating weights $w(r_i)$ for $r_i = 1 \cdots m$ using the current estimate of $\Delta\mathbf{d}$ and solving Equation 5 to estimate a new $\Delta\mathbf{d}$ with the weights fixed. Among many proposed shapes for $w(\cdot)$, we use [11]

$$w(r_i) = \begin{cases} 1 & |r_i| < T \\ \frac{T}{|r_i|} & |r_i| > T \end{cases} \tag{6}$$

where T is a threshold that can be tuned. A small T will treat many samples as outliers. With the addition of the weight function, Equation 5 becomes

$$(\mathbf{w}{\mathbf{I}_2'}^2 + \alpha\mathbf{D} + \alpha_2\hat{\mathbf{I}})\Delta\mathbf{d} = \mathbf{w}\mathbf{I}_2'\mathbf{e} - (\alpha_1\mathbf{D} + \alpha_2\hat{\mathbf{I}})\mathbf{d} + \alpha_2\mathbf{d}^{t.p.} + \mathbf{b} \tag{7}$$

where $\mathbf{w} = diag(w(r_1)\cdots w(r_m))$. All of the results presented in this work are obtained with one iteration of the above equation unless otherwise specified. Current implementation of the AM algorithm with the IRLS takes 0.015s to generate a dense displacement field of size 1300×60 on a 3.4GHz P4 CPU(not including the DP run time). The computation time increases linearly with the size of images.

Frame Selection. The ultrasound probe is tracked in navigation/guidance systems to provide spatial information, to generate freehand 3D ultrasound, or to facilitate multi-modality registration. Through a calibration process, the 6DOF motion of the probe in the sensor coordinate system is transformed into image coordinate system [12]. The mean of the absolute motion value of all pixels in 3D, $\langle|v_x|\rangle$, $\langle|v_y|\rangle$ and $\langle|v_z|\rangle$, can be analytically related to the 6DOF sensor readings using straightforward and efficient geometric computations. For frame i and j to be selected from a sequence of frames for elastography,

$$Q_{i,i} = k_x\ \langle|v_x|\rangle^2 + k_y\ \langle|v_y|\rangle^2 + k_z\frac{\|\ \langle|v_z|\rangle - v_{z,opt}\|^3}{\langle|v_z|\rangle + c} \tag{8}$$

should be minimized where k_x, k_y, and k_z are weights for lateral, out-of-plane and axial displacements and $v_{z,opt}$ is the optimum axial motion. Please refer to [12] for a rationale of the shape this function. Note that the selected pairs are not necessarily consecutive frames. The parameters, k_x, k_y, k_z, $v_{z,opt}$ and c are manually tuned to 1, 2, 1, 0.7 and 1 for the AM elastography method.

3 Simulation, Phantom and Patient Results

Simulation Results. RF ultrasound data of two phantoms are simulated using Field II [13]. The first phantom is $50 \times 10 \times 55$mm and the second one is $36 \times 10 \times 25$mm. They are both made of homogeneous and isotropic material: the first one is uniform and the second one contains a circular hole filled with water, simulating a blood vessel in tissue (Figure 2 (a)). A uniform compression in the z direction is applied and the 3D displacement field of the phantom is calculated using ABAQUS finite element package (Providence, RI). The Poisson's ratio is set to $\nu = 0.49$ in both phantoms to mimic real tissue, which causes the phantom to deform in x & y directions as a result of the compression in the z direction.

Respectively 5×10^5 and 1.4×10^5 scatterers with uniform scattering strengths are uniformly distributed in the first and second phantom, ensuring more than 10 scatterers exist in a resolution cell. The scatterers are distributed in the 8mm diameter vein also (Figure 2 (a)). To construct deformed ultrasound images, the displacement of all of the scatterers is calculated by interpolating the displacement of the neighboring nodes in the finite element analysis. The parameters of the probe are set to mimic Siemens 5-10MHz probes. The probe frequency is 7.27MHz, the sampling rate is 40MHz and the fractional bandwidth is 60%.

The first phantom undergoes uniform compressions in the z direction to achieve strain levels of 2% to 14% in 2% intervals. Ground truth integer displacement values are used as the initial estimate for AM to decouple the performance of DP from AM. Accurate subpixel displacement field is calculated with AM and the mean strain values are compared with the ground truth (Figure 1 (a)-(c)). The results are only shown for 2%, 4%, 8% and 14% compression for better visualization. The results with two threshold values for IRLS and without IRLS demonstrate that outlier rejection does not affect the mean strain value, while increasing the regularization weight α_a increases underestimation of the displacement. The rate of increase of the underestimation with increasing α_a is significantly more with the unbiased regularization (dashed line) as expected.

Significantly higher signal to noise ratio (SNR) [2] values can be achieved with outlier rejection (Figure 1 (d)-(f)) without over-smoothing the image with high α_a values. To show the performance of the overall method, the initial integer displacement field is calculated with DP and accurate displacement field is calculated with (Figure 1 (g)-(i)). The SNR values are less than previous case especially at high strain values, where DP results deviates from ground truth.

The second simulation experiment is designed to show the effect of smoothness weight and IRLS threshold on contrast to noise ratio (CNR) [2] when the correlation is lower in parts of the image due to fluid motion. The phantom

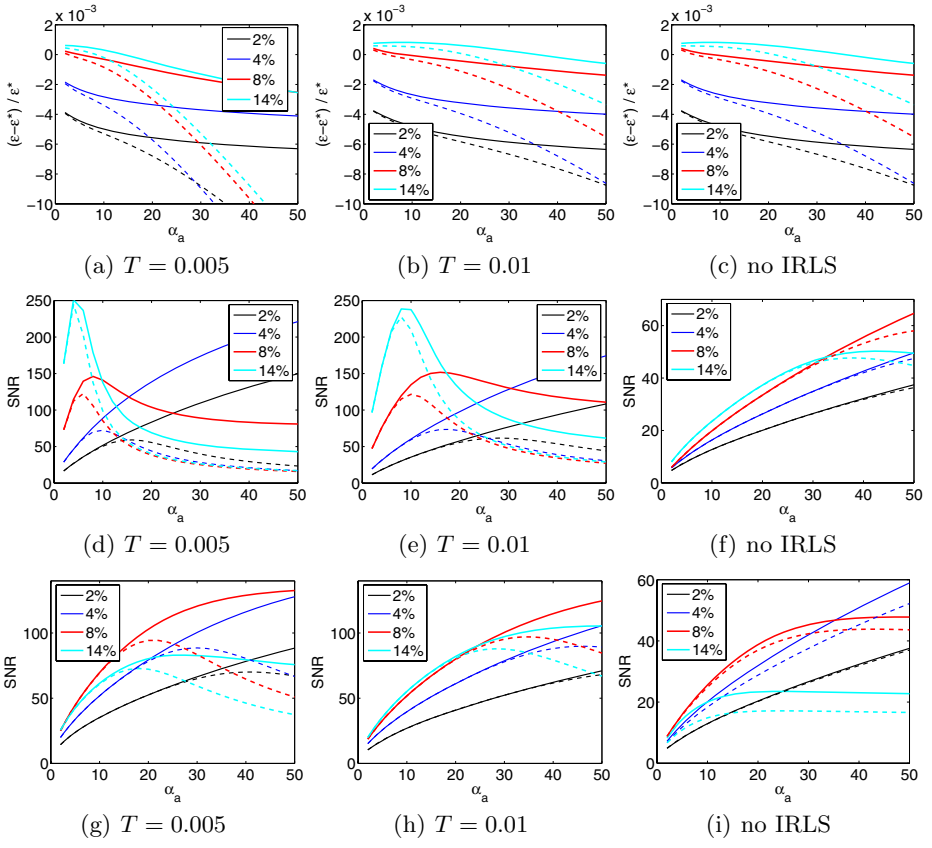
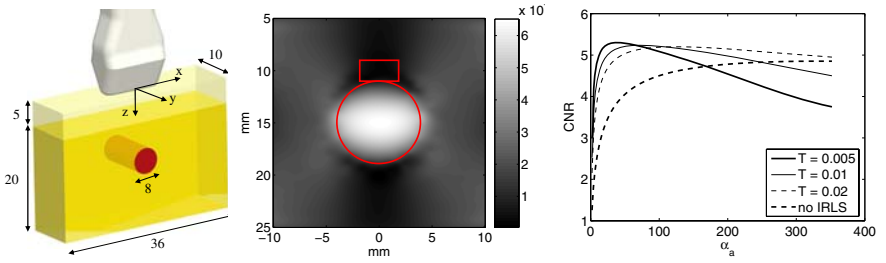


Fig. 1. Mean and SNR of the elastograms of the Field II simulated uniform phantom at four different compression levels (shown in percentage) for three IRLS T values. The solid and dashed lines correspond to biased and unbiased regularizations respectively. (a)-(c) shows the relative underestimation of the strain. ϵ is the mean strain calculated with the elastography method and ϵ^* is the ground truth. (d)-(f) shows the SNR of the AM. (g)-(i) shows the SNR of the AM with initial displacements found by DP.

contains a vein oriented perpendicular to the image plane (Figure 2). The initial integer displacement is generated with DP. The background window for CNR calculation is located close to the target window to show how fast the strain is allowed to vary, a property related to the spatial resolution. The maximum CNR with IRLS is 5.3 generated at $T = 0.005$ and $\alpha_a = 38$, and without IRLS is 4.8 at $\alpha_a = 338$. Such high α_a value makes the share of the data term in the cost function very small and causes over-smoothing.

Phantom Results. We perform freehand palpation experiment on a breast phantom to examine the performance of the frame selection technique. 50 frames of RF ultrasound data are acquired using a Siemens Antares system (Issaquah, WA). Our custom data acquisition program is connected to the Axius Direct Research Interface to send the command for capturing RF data. At the same



(a) simulation phantom (b) finite element strain (c) CNR

Fig. 2. The target (circular) and background (rectangular) windows for CNR calculation of (c) are shown in (b)

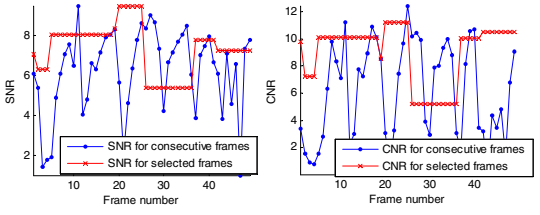
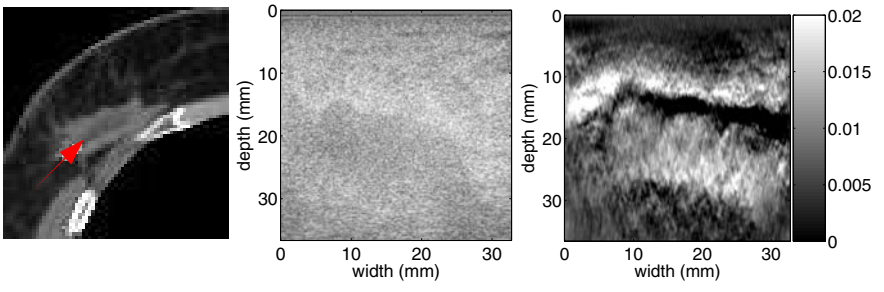


Fig. 3. The SNR and CNR of the phantom experiment with and without frame selection



(a) CT image (b) B-mode image (c) strain image

Fig. 4. Patient experiment results. The arrow points to the lumpectomy cavity.

time, the program collects tracking information from a Polaris tracker (Waterloo, Canada). Currently, the RF frames are stored on the ultrasound system and are processed offline. Figure 3 shows the SNR and CNR results. In automatic frame selection, $Q_{i,j}$ (equation 8) for any two frames i, j in a buffer of size 15 frames is calculated. For the two frames which give the minimum Q, the strain image is obtained. The next image is then fed to the buffer, its first image is removed and the frame selection is performed again. The automatic frame selection gives 8 frame pairs for strain calculation (as seen in the figure by 8 SNR and CNR

values). Without frame selection, 49 strain images are calculated. The average CNR and SNR values are improved from 4.91 to 7.19 and from 4.98 to 5.88 with frame selection.

Patient Results. We have acquired freehand palpation ultrasound RF data using the Siemens Antares system from patients approximately four weeks after lumpectomy. The ultrasound probe is tracked with the Polaris tracking system. Optimal frame selection is performed to select images for elastography using the AM method. The strain image (Figure 4) shows that the AM method can detect the thin hard scar tissue even though it is close to the cavity fluids which undergo incoherent motions and cause signal decorrelation. Since the AM method finds the displacement of all the samples on an A-line at the same time, the correlated data at the top and bottom of the cavity guide the method to find the correct displacement inside the cavity where the data is decorrelated.

4 Discussion and Conclusion

We introduced a novel method for calculating a dense displacement map by analytic minimization of a cost function. We used the IRLS method from robust statistics to make the tracking resistant to outliers. Moreover, we exploited the tracking data to optimize frame selection. Through simulation studies using Field II and finite element analysis, we showed that the proposed AM method generates high quality displacement estimates. The elastography method works in real-time. A comparison of the IRLS method with quality guided displacement tracking [14] which also aims for robustness is a subject of future work.

We chose the novel application of the lumpectomy cavity localization as the hard scar tissue is relatively thin and demands a high resolution elastography method. Also, incoherent fluid motions in the cavity causes large decorrelations, requiring a robust method. We have an active Institutional Review Board(IRB) protocol and have promising results from 9 patients which will be published in future work.

Acknowledgments. We would like to thank Shelby Brunke for ultrasound support. Hassan Rivaz is supported by DoD Breast Cancer Research Predoctoral Award and by Link Foundation Fellowship. This work is funded by Breast Cancer Research Foundation, by Siemens Medical Solutions and by CISST ERC NSF EEC-9731748.

References

1. Offersen, B., Overgaard, M., Kroman, N., Overgaard, J.: Accelerated partial breast irradiation as part of breast conserving therapy of early breast carcinoma: a systematic review. Radiotherapy and Oncology 90(1), 1–13 (2009)
2. Ophir, J., Alam, S., Garra, B., Kallel, F., Konofagou, E., Krouskop, T., Varghese, T.: Elastography. Annu. Rev. Biomed. Eng. 213, 203–233 (1999)

3. Hall, T., Zhu, Y., Spalding, C.: In vivo real-time freehand palpation imaging. Ultrasound Med. Biol. 29, 427–435 (2003)
4. Zahiri, R., Salcudean, S.: Motion estimation in ultrasound images using time domain cross correlation with prior estimates. Biomed. Eng. 53(10), 1990–2000 (2006)
5. Pellot-Barakat, C., Frouin, F., Insana, M., Herment, A.: Ultrasound elastography based on multiscale estimations of regularized displacement fields. IEEE Trans. Med. Imag. 23(2), 153–163 (2004)
6. Brusseau, E., Kybic, J., Deprez, J., Basset, O.: 2-d locally regularized tissue strain estimation from radio-frequency ultrasound images: Theoretical developements and results on experimental data. IEEE Trans. Med. Imag. 27(2), 145–160 (2008)
7. Rivaz, H., Boctor, E., Foroughi, P., Fichtinger, G., Hager, G.: Ultrasound elastography: a dynamic programing approach. IEEE Trans. Med. Imag 27, 1373–1377 (2008)
8. Lyshchik, A., Higashi, T., Asato, R., et al.: Thyroid gland tumor diagnosis at US elastography. Radiology 237, 202–211 (2005)
9. Kadour, M., Noble, A.: Assisted-freehand ultrasound elasticity imaging. IEEE Trans. Ultrason. Ferroelectr. Freq. Control. 129, 36–43 (2009)
10. Lindop, J., Treece, G., Gee, A., Prager, R.: An intelligent interface for freehand strain imaging. Ultrasound Med. and Biol. 34(7), 1117–1128 (2008)
11. Stewart, C.: Robust parameter estimation in vision. SIAM 41(3), 513–537 (1999)
12. Foroughi, P., Hager, G., Boctor, E.: Robust elasticity imaging using external tracker. In: IEEE Int. Symp. Biomed. Imag. (in press, 2009)
13. Jensen, A.: Field: A program for simulating ultrasound systems. Medical & Biological Engineering & Computing 34, 351–353 (1996)
14. Chen, L., Treece, G., Lindop, J., Gee, A., Prager, R.: A quality-guided displacement tracking algorithm ultrasonic elasticity imaging. Med. Imag. A. 13, 286–296 (2009)

Image Guidance for Spinal Facet Injections Using Tracked Ultrasound

John Moore[1], Colin Clarke[2], Daniel Bainbridge[2,3], Chris Wedlake[1], Andrew Wiles[1,4], Danielle Pace[1,5], and Terry Peters[1,4,5]

[1] Imaging Research Laboratory, Robarts Research Institute, London, Canada
jmoore@imaging.robarts.ca
[2] Division of Anesthesia, The University of Western Ontario, Canada
[3] Canadian Surgical Technologies and Advanced Robotics (CSTAR), Canada
[4] Biomedical Engineering Graduate Program, The University of Western Ontario, Canada
[5] Department of Medical Imaging, Department of Medical Biophysics, The University of Western Ontario, Canada

Abstract. Anesthetic nerve blocks are a common therapy performed in hospitals around the world to alleviate acute and chronic pain. Tracking systems have shown considerable promise in other forms of therapy, but little has been done to apply this technology in the field of anesthesia. We are developing a guidance system for combining tracked needles with non-invasive ultrasound (US) and patient-specific geometric models. In experiments with phantoms two augmented reality (AR) guidance systems were compared to the exclusive use of US for lumbar facet injection therapy. Anesthetists and anesthesia residents were able to place needles within $0.57mm$ of the intended targets using our AR systems compared to $5.77mm$ using US alone. A preliminary cadaver study demonstrated the system was able to accurately place radio opaque dye on targets. The combination of real time US with tracked tools and AR guidance has the potential to replace CT and fluoroscopic guidance, thus reducing radiation dose to patients and clinicians, as well as reducing health care costs.

1 Introduction

A wide variety of pain conditions are treated with needle injection therapies. Selective nerve blocks require careful placement of the injection needle, both to ensure effective therapy delivery and to avoid damaging neighboring tissues. The facet joint is a source of chronic pain for between 15% to 45% of patients with chronic lower back pain[1]. Lumbar facet joint injections are a good example of a particularly challenging therapy; the small, narrow channel between vertebrae, the oblique entry angle, relatively deep location and proximity to nerve tissue can make accurate treatment delivery quite difficult.

Most facet injections are performed using fluoroscopy or CT guidance, which involves delivering radiation dose to the patient and health care providers. Recently,

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 516–523, 2009.

a number of groups have been developing guidance techniques with the goal of reducing or eliminating this radiation dose while either maintaining or improving patient safety. For example, Proschek *et al* [2] have developed a guidance system that uses a laser projection system to limit the radiation dose required for needle delivery. Magnetic tracking systems (MTS) have shown promise for improving the guidance of a wide variety of interventional procedures [3]. Tracking technology is relatively inexpensive compared to 3D US systems or fluoroscopy, making it a viable option for any clinical setting. Bruners *et al* have demonstrated a system that combines interventional CT with tracked needles to facilitate lumbar facet needle guidance [4], while Galiano *et al* have shown that ultrasound (US) can be used instead of fluoroscopy or CT for lumbar facet joint injections [5]. US technology is both ubiquitous and considerably less expensive than fluoroscopy or CT, and eliminates radiation dose to both patient and clinician. However, on its own US can be difficult to interpret and the high frequencies required to adequately visualize small needles will not always penetrate deeply enough to visualize the target facet joint.

By integrating 2D US with virtual representations of anatomical targets and the tracked therapy delivery needle, it is possible to greatly enhance the US information available to the clinician. The overall goal of the present work is to develop an anesthesia delivery system that allows the integration of virtual models derived from diagnostic imaging such as CT and tracked tools, all intended to enhance real time US data. Our specific goal in the present work is to assess the accuracy of augmented reality (AR) assistance for US guided lumbar spinal facet injections. We evaluate two information delivery strategies: a simple system representing only the tools being used (US and needle), and a more elaborate system that also incorporates a patient specific spine model for additional guidance.

2 Materials and Methods

Four anesthetists and four anesthesia residents evaluated three different guidance systems (see Figure 1) to place a tracked needle on target for facet therapy injections in a lumbar spine phantom:

GS_{us} GS_{tools+} GS_{model+}

Fig. 1. The three guidance systems evaluated. The virtual needle tool also has x,y and z axis extensions with $10mm$ markers to help locate the tip and its trajectory.

- US only (GS_{us}): the users were restricted to using the US machine monitor. They were free to manipulate the location and angle for the US transducer;
- US plus virtual tool representations (GS_{tools+}): the users had full use of ultrasound, plus representations of the US and needle tool: the AR display consisted of the normal fixed view of the US fan with 'top' and 'side' views showing the angle and tip of the needle relative to the US beam;
- Complete 3D AR system (GS_{model+}): the users had full use of the US, representations of US transducer and needle, plus an anatomical model of the spine based on a high resolution CT. The user had full control of their view of the scene.

The GS_{tools+} system is intended for applications where no anatomical models are available (e.g., no pre-procedure imaging is done, as is the case for most peripheral nerve blocks). It was expected this system would have an advantage over US guidance since it allows the anesthetist to localize the needle tip even when it is not entirely in plane in the US. This system will be evaluated specifically for use in peripheral nerve block applications in future studies.

To date we have performed a single cadaver study to assess the logistics of translating our techniques to a clinical environment.

Apparatus. Our phantom experiments were carried out using a plastic lumbar spine that was rigidly fixed inside a lexan box using four high grade stainless steel screws to hold each vertebra (see Figure 2). A total of 17 concave landmark divots were machined randomly into the outside of the container for registration purposes. The spine was embedded in a tissue mimicking polyvinyl alcohol cryogel (PVA-C) material (10% PVA, two freeze-thaw cycles) [6]. To remove needle tracks left behind in the phantom after repeated experiments, the PVA-C was heated above 60° Celcius to melt it into a liquid state after which the freeze-thaw cycles were repeated to return the PVA to its solid tissue-mimicking state. The phantom was imaged using a 64 slice LightSpeed VCT scanner (General Electric, Milwaukee, WI, USA), ($0.441mm$ by $0.441mm$ in plane, $0.625mm$ slice thickness), to generate an accurate surface model of the spine. An expert identified the ideal target facet locations in the phantom CT by manually segmenting each facet joint using ITK-SNAP (www.itksnap.org), and a spine model was generated using the Marching Cubes algorithm in VTK (Kitware Inc.), as shown in Figure 2. Landmark divot locations were defined in the CT image to register the tracked US and needle into the CT/model/target coordinate frame. Imaging was provided by an Aloka US system with a 5 MHz neuro transducer.

Measuring Accuracy. For the actual therapy delivery, a magnetic tracking system is required since the delivery needle is capable of bending inside tissue. The Aurora MTS (NDI, Waterloo, ON, Canada) was used to track needles and US. The needle (stylus and cannula) used was a 5 degrees of freedom (DOF) 18 gauge, 200 mm long needle (Figure 2) with the sensor located at its tip. 6DOF sensors were used to track the US transducer and to provide a reference for the phantom itself. The US transducer was calibrated using a Z-bar phantom

(a) (b) (c)

Fig. 2. Phantom, tools and targets: (a) spine inside the phantom, (b) phantom with tissue-mimicking PVA-C, tracked needle tool, US transducer and tracker field generator (on right), (c) Facet targets in context of the spine model generated from CT data

[7]. Since the MTS has an inherent coordinate frame based on the magnetic field generator, all tool coordinate vectors ($\overline{p}_{fg}$) are relative to this frame of reference. This could cause problems if the phantom or field generator move relative to each other during an experiment. Consequently the reference tool was rigidly attached to the spine phantom box to yield coordinates relative to the phantom, thus ensuring tracked tool coordinates ($\overline{p}_{mts}$) remained consistent with the phantom coordinate frame:

$$\overline{p}_{mts} = T_{ref}^{-1}\left(\overline{p}_{fg}\right), \tag{1}$$

where T_{ref} represents the transform of the reference tool. In clinical applications, placing this reference tool near the target will also help compensate for patient shifting and breathing motion. The tracking coordinate frame was registered to the CT-generated model using a tracked sphere-tipped tool that mated with the divots on the outside of the phantom box, generating a very accurate point-based registration. Using this point based registration transform ($T_{phantom}$), we translated the measured needle coordinates ($\overline{p}_{mts}$) into the CT coordinate frame:

$$\overline{p}_{meas} = T_{phantom}\left(\overline{p}_{mts}\right), \tag{2}$$

where $\overline{p}_{meas}$ is the measured needle tool tip location CT space. Trueness is then the distance defined by the magnitude of the error vector,

$$d_{err} = |\overline{p}_{error}| = \left|\overline{p}_{meas} - \overline{p}_{target}\right|, \tag{3}$$

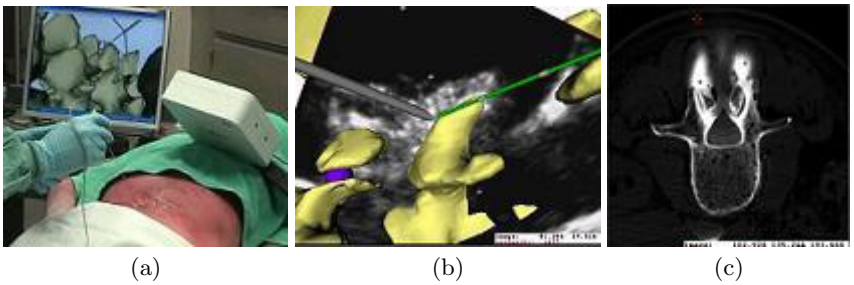
where $\overline{p}_{target}$ is the closest point on the facet joint polydata as defined in the CT. When $\overline{p}_{meas}$ was located inside the target polydata, the error was zero.

Software. The GS_{model+} system used a software platform that presents the real-time tracked US and tools in a 3D context, allowing the user to interactively select which data to display: the pre-procedural CT, real time US image, in addition to geometric models, from any desired point of view [8].

With the GS_{model+} system, the user was able to toggle between two "standard views" of their own choosing. Typically, this feature allowed the user to flip between orthogonal views to provide accurate perception of needle trajectory and needle depth. The GS_{tools+} system was based on the same software platform.

Implementation. Each subject gained familiarity with the AR guidance systems using a training apparatus consisting of tracked ultrasound and needle and a series of small targets in a water bath. Surgical targets were shown to the subjects prior to performing the experiment but were not available during the experiment, since these data would not be available during an actual needle therapy delivery. After the subjects were comfortable with the systems, they proceeded to place the delivery needle at the target locations. A total of six target locations were used: facet joints between L5 and L4, L4 and L3, L3 and L2. Each guidance modality was used on each target, yielding 18 deliveries for each subject, six for each guidance system. The target order and modality were uniquely randomized for each subject in order to minimize any bias.

Cadaver Study. A proof of concept cadaver study was performed using the GS_{model+} guidance system to assess the system in a clinical setting. This study was performed according to the University of Western Ontario requirements for the use of cadavers. Seven homologous points representing the spinal and lateral processes in the region of interest were defined in the pre-procedural CT and interactively in the OR using 2D US to perform the point based registration. This procedure took approximately five minutes. To properly mimic a clinical therapy, radio opaque dye was injected when the user believed the needle was in the correct position. An anesthesia resident performed the needle delivery using the GS_{model+} guidance system, placing radio opaque dye at the left and right L2-3 and L3-4 facet joints. Accuracy was assessed independently by a radiologist using a post-procedure CT (see Figure 3).



(a) (b) (c)

Fig. 3. Cadaver study and results. (a) OR environment showing AR scene, delivery needle and MTS field generator, (b) close-up of GS_{model+} system including needle (gray), US fan and projected y axis (green), (c) post-procedure CT showing radio opaque dye at the L2-3 facet joints.

Table 1. RMS distance errors in mm for each image guidance system. 'Correct target' excludes outliers where the subject identified the wrong facet joint but accurately reached the (incorrect) target.

	GS_{us}	GS_{tools+}	GS_{model+}
all data	10.22	8.45	0.57
correct target only	5.77	4.87	0.57

3 Results and Discussion

The phantom experiment was a $3 \times 6 \times 8$ factorial design, the factors being guidance system, target and user, respectively. An analysis of variance (ANOVA) was performed to test for significant variance across the three factors. When a significant difference was found in the ANOVA, Tukey's Honestly Significant Difference test was applied to determine which levels of the given factor were different. No significant difference was noted across users or targets ($p > 0.05$).

The use of GS_{model+} showed a marked superiority to GS_{US} and GS_{tools+}. One curious phenomena was noted in several subjects: on six occasions using the GS_{tools+} system and five occasions with the GS_{US} system, users managed to correctly place the needle tip inside a facet joint, but selected the incorrect joint. This problem never arose with the GS_{model+} system. Table 1 displays root mean squared (RMS) distance error data considering all trials, including cases where the anesthetist placed the needle tip "correctly" in the wrong facet joint, as well as the distance error data excluding these cases. In both cases, there is a significant difference ($p << 0.05$) between the GS_{model+} system and the other two systems. How common this form of error is in practice is unclear, however the use of an AR system incorporating anatomical models would appear to circumvent the problem.

Figure 4 provides a graphical representation of therapy delivery locations relative to targets. Overall, the GS_{tools+} system showed slightly better RMS error than US alone, but not at a statistically significant level. The cognitive demands of this new guidance system may have contributed to these results, as several clinicians commented that the computer monitor was rather "busy". Larger sample sizes are planned for future research which should help to quantify the advantage of the GS_{tools+} system, particularly for peripheral nerve block applications where no pre-procedural imaging is done.

An unpaired t-test was performed to compare residents and anesthetist accuracy, and no significant difference was found. The anesthesia residents involved in this study had significant experience with US guided peripheral nerve blocks, so this lack of difference between novices and experts is not surprising. Only two of the participants had prior experience with facet joint injections. Several of the participants commented on the training and educational potential of the guidance systems, both for identifying anatomical landmarks (using the GS_{model+} system) and recognizing when the needle tip is in or out of the US plane (GS_{tools+} system). Anecdotally, all users described the GS_{model+}

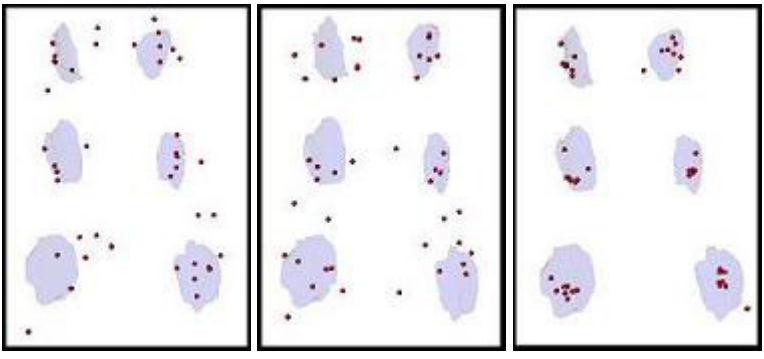


Fig. 4. Distribution of needle placements in coronal view: blue clouds are the facet joint targets, spheres are needle delivery locations. (a) US alone, (b) GS_{tools+} system, (c) GS_{model+} system.

guidance system as a substantial improvement over existing systems (US alone and fluoroscopy).

Introducing the tracking system into the OR environment for the cadaver study posed no significant difficulties. Attention was paid to minimize the presence of magnetic materials in the vicinity of the MTS field generator. Although we did not record the number of needle penetrations necessary for accurate placement, the clinician commented that the new system greatly reduced the number of penetrations and the time required to position the needle correctly. The x, y, and z axes projected from the virtual needle tip were especially useful for determining trajectory and tip location. The radiologist report stated signal enhancement (due to injected radio opaque dye) in and surrounding facet joints right and left, L2-3 and L3-4. This correlates with a clinically significant injection (see Figure 3 (c)).

4 Conclusions and Future Work

Our results indicate that AR assisted ultrasound guidance for facet injections shows tremendous promise for increasing patient safety and comfort, greatly reducing or even eliminating radiation dose to both patient and clinicians, as well as reducing health care costs. The use of tracked US and needles in an AR environment also has a great deal of potential for application in nerve block clinics and anesthesia training.

Our next steps include the application and assessment of several features, including a more clinically relevant CT to US registration algorithm. We are also evaluating the use of a volume rendered display of 3D reconstructed ultrasound data, since this would eliminate the need for registering pre-procedural CT data and models. The need to compensate for respiratory motion and changes in position during the clinical intervention may be reduced by the use of a tracked

reference tool in the vicinity of the target tissues, though this will need further study. An assessment of the cognitive demands presented by the AR systems will also help to properly interpret their significance. Educational and clinical applications for peripheral nerve blocks are also in the planning stages.

Acknowledgments. We would like to thank the Department of Anesthesia, London Health Sciences Center, Canada for consultations and assistance with experiments. We would also like to thank Dr. Gord Campbell (National Research Council, London Canada) for assistance with PVA-C procedures and resources, Dr. Tim Wilson and Mahar Sabalbal (University of Western Ontario, London Canada) for assistance with geometric models used in the cadaver study, and CT technologists at LHSC- University Hospital, London, Ontario for CT data acquisition.

References

1. Boswell, M., Colson, J., Sehgal, N., Dunbar, E., Epter, R.: A systematic review of therapeutic facet joint interventions in chronic spinal pain. Pain Physician 10, 229–253 (2007)
2. Proschek, D., Kafchitsas, K., Rauschmann, M., Kurth, A., Vogl, T., Geiger, F.: Reduction of radiation dose during facet joint injection using the new image guidance system sabersource: a prospective study in 60 patients. European spine journal 18, 546–553 (2009)
3. Wood, B., Zhang, H., Durrani, A., Glossop, N., Ranjan, S., Lindisch, D., Levy, E., Banovac, F., Borgert, J., Krueger, S., Kruecker, J., Viswanathan, A., Cleary, K.: Navigation with electromagnetic tracking for interventional radiology procedures:A feasibility study. Journal of vascular and interventional radiology 16, 493–505 (2005)
4. Bruners, P., Penzkofer, T., Nagel, M., Elfring, R., Gronloh, N., Schmitz-Rode, T., Gunther, R., Mahnken, A.: Electromagnetic tracking for CT-guided spine interventions: phantom, ex-vivo and in-vivo results. European Radiology 19, 990–994 (2009)
5. Galiano, K., Obwegeser, A.A., Walch, C., Schatzer, R., Ploner, F., Gruber, H.: Ultrasound-guided versus computed tomography-controlled facet joint injections in the lumbar spine: A prospective randomized clinical trial. Regional Anesthesia and Pain Medicine 32(4), 317–322 (2007)
6. Surry, K., Austin, H., Fenster, A., Peters, T.: Poly(vinyl alcohol) cryogel phantoms for use in ultrasound and MR imaging. Physics in Medicine and Biology 49(24), 5529–5546 (2004)
7. Gobbi, D., Comeau, R., Peters, T.: Ultrasound probe tracking for real-time ultrasound/MRI overlay and visualization of brain shift. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 920–927. Springer, Heidelberg (1999)
8. Moore, J., Guiraudon, G., Jones, D., Hill, N., Wiles, A., Bainbridge, D., Wedlake, C., Peters, T.: 2D ultrasound augmented by virtual tools for guidance of interventional procedures. Studies in Health Technology and Informatics, 322–327 (2007)

Cervical Vertebrae Tracking in Video-Fluoroscopy Using the Normalized Gradient Field

Rianne Reinartz[1], Bram Platel[2], Toon Boselie[3], Henk van Mameren[4], Henk van Santbrink[3], and Bart ter Haar Romeny[1]

[1] Dept. of Biomed. Eng., Eindhoven University of Technology, The Netherlands
[2] Dept. of Biomed. Eng., Maastricht University Medical Centre, The Netherlands
[3] Dept. of Neurosurgery, Maastricht University Medical Centre, The Netherlands
[4] Dept. of Epidemiology, Maastricht University Medical Centre, The Netherlands

Abstract. For patients with neck problems valuable functional and diagnostic information can be obtained from a fluoroscopy video of a flexion-extension movement of the cervical spine. In most cases physicians have to manually extract the vertebrae, making the analysis of these video sequences tedious and time consuming. In this paper we propose an automatic fast and precise method for tracking cervical vertebrae. Our method relies only on a rough selection of template areas of each vertebra in a single frame of the video sequence. Compared to existing automated methods, no contours need to be extracted and no vertebra segmentation is required. Tracking is done with a normalized gradient field, using only the gradient orientations as features. Experimental results show that the algorithm is robust and able to track the vertebrae accurately even if they are partially occluded or if a disc prosthesis is present.

1 Introduction

The order in which the segments of the cervical spine contribute to the motion of the skull with respect to the thorax during anteflexion and retroflexion (forward and backward bending of the neck), has proven to provide valuable functional and diagnostic information for patients with neck complaints [1,2]. To analyze the movement of the cervical spine in detail, a video-fluoroscopy recording (i.e. an x-ray video), is created for a movement from maximum extension to maximum flexion of the neck (Fig. 1). From this recording the rotation and translation of the cervical vertebrae with respect to each other need to be extracted.

Automatic spinal measurement techniques have been studied for the lumbar as well as the cervical spine. Generally these methods are either based on landmark detection, or on the extraction of vertebral boundaries. Due to noise, occlusion and unclear vertebral contours, the detection of landmarks or contours is difficult. Techniques exploiting the Hough transform [3,4], active appearance models [5], level sets [6] and polar signatures [7], have been used to find the unclear vertebral

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 524–531, 2009.

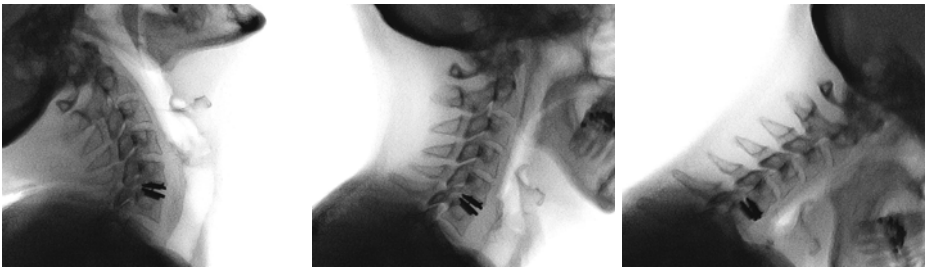


Fig. 1. Three frames of a typical cervical fluoroscopy video. The images respectively show the patient in retroflexion, neutral position and anteflexion.

boundaries and track them in fluoroscopy videos. These methods often use large sets of templates to capture the great variability in the shape of the vertebrae, require extensive manual interaction, are computationally heavy and are complex to implement.

For the motion analysis of the cervical spine, exact contour extraction and vertebra segmentation is not necessary. We therefore propose a method based on the normalized gradient field that can automatically track cervical vertebra in a video-fluoroscopy sequence, requiring user interaction in only a single frame of the video. This eliminates the cumbersome and computationally heavy segmentation of the vertebral contours.

2 Methods

2.1 Cervical Video-Fluoroscopy

To acquire the cervical fluoroscopy video, the subject is placed on a stool, with his shoulders perpendicular to the image intensifier. The shoulder nearest to the image intensifier is placed directly against it. The subject is asked to move his head in maximal extension without moving the upper part of his body. As soon as the recording is started, the subject is instructed to move his head in the sagittal plane from maximal extension to maximal flexion, without moving the upper part of the body. The subjected is asked to perform the full movement in about 15 seconds. It is important that the subject's shoulders are kept as low as possible while making the cervical fluoroscopy videos to ensure that all the cervical vertebrae are visible. The movement of the cervical spine should be as fluent as possible to prevent for sudden large rotations and translations between consecutive frames. The fluoroscopic recordings were made with a digital X-ray detector, capturing frames of 1024×1024 pixels, at 10 frames per second. The recordings were stored without compression.

2.2 Template Selection

The tracking algorithm requires a template of each vertebral body and the skull, that can be tracked throughout the whole video sequence. Due to partial

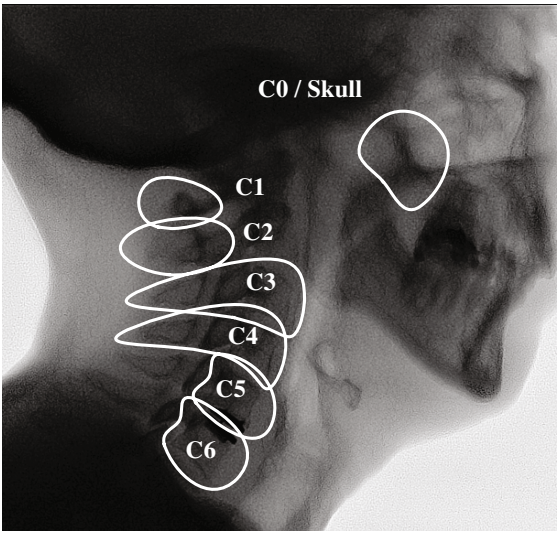


Fig. 2. An example set of templates for the vertebrae and the skull

occlusion by the shoulder of C5 and C6 only the anterior parts of these vertebrae are visible during retroflexion, which is why these anterior parts are taken as a template. The anterior parts of C1 and C2 appear as one, as the dens of C2 moves into C1 to form the atlanto-axial joint. This is why only the posterior parts of these vertebrae are used as a template. The bottom part of the maxillary sinus and the palatum durum are projected in the x-ray image as a stable and structure rich area of the skull. This is therefore an ideal template to track the movement of C0 (the skull). C7 is usually completely occluded by the shoulder and can therefore not be tracked. Figure 2 shows the template regions. The user is asked to draw these templates in a single frame of the fluoroscopy video. The templates are selected by positioning a closed spline over the template area. This spline is positioned by moving a small number of points (the user can choose the amount, but these are usually 4 points) over the image.

2.3 Vertebra Tracking

We define a user selected portion of a vertebral body as the *template* T, and the subsequent image in the fluoroscopy video sequence as the *reference* R. The goal is to find a rigid transformation φ such that the "distance" between the template T, and the reference R, is as small as possible. The rigid transformation is defined as:

$$\varphi(\boldsymbol{\gamma}, \boldsymbol{p}) = \begin{pmatrix} \cos(\gamma_1) & -\sin(\gamma_1) \\ \sin(\gamma_1) & \cos(\gamma_1) \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} + \begin{pmatrix} \gamma_2 \\ \gamma_3 \end{pmatrix}. \tag{1}$$

Given a distance measure $\mathcal{D}$, the goal is to find a minimizer $\boldsymbol{\gamma}$ of

$$f(\boldsymbol{\gamma}) := \mathcal{D}\Big(R(\cdot), T(\varphi(\boldsymbol{\gamma}, \cdot))\Big). \tag{2}$$

The cervical vertebrae tracking is done on a video sequence of fluoroscopic images. Partial occlusion of the vertebrae by other bones causes significant intensity variations. A proper distance measure to register the reference and target should therefore not depend on the intensities of the image. We have chosen to use the normalized gradient field (NGF) as described by Haber [8]. The NGF contains the local orientations of the image gradient. In regions where the intensities are nearly constant the gradient orientation is unreliable. To lower the weight of these gradient orientations a parameter ϵ is introduced so that the normalized gradient $\hat{\nabla}_\epsilon$ becomes

$$\hat{\nabla}_\epsilon I := \frac{\nabla I}{\sqrt{|\nabla I|^2 + \epsilon^2}}, \quad \text{with} \quad \epsilon = \frac{\eta}{A} \int_\Omega \mid \nabla I(\boldsymbol{p}) \mid d\boldsymbol{p}, \tag{3}$$

where η is the estimated noise level in the image and A is the area of the image domain Ω. With $S := T \cdot \varphi$, the normalized gradient vectors at position $\boldsymbol{p}$ become $\{\hat{R}_x(\boldsymbol{p}), \hat{R}_y(\boldsymbol{p})\} = \hat{\nabla}_\epsilon R(\boldsymbol{p})$ and $\{\hat{S}_x(\boldsymbol{p}), \hat{S}_y(\boldsymbol{p})\} = \hat{\nabla}_\epsilon S(\boldsymbol{p})$. As the gradient fields are approximately normalized, the angle $\theta(\boldsymbol{p})$ between the vectors $\{\hat{R}_x(\boldsymbol{p}), \hat{R}_y(\boldsymbol{p})\}$ and $\{\hat{S}_x(\boldsymbol{p}), \hat{S}_y(\boldsymbol{p})\}$ is given by:

$$\begin{aligned} \left|\hat{R}_x(\boldsymbol{p})\hat{S}_y(\boldsymbol{p}) - \hat{R}_y(\boldsymbol{p})\hat{S}_x(\boldsymbol{p})\right| &= \left|\{\hat{R}_x(\boldsymbol{p}), \hat{R}_y(\boldsymbol{p})\}\right| \ \left|\{\hat{S}_x(\boldsymbol{p}), \hat{S}_y(\boldsymbol{p})\}\right| \cdot \sin(\theta(\boldsymbol{p})) \\ &\approx \sin(\theta(\boldsymbol{p})). \end{aligned} \tag{4}$$

As we want to minimize the differences in the angles between the normalized gradient field of reference R and template T we define the distance between two normalized gradient vectors as

$$d(R(\boldsymbol{p}), T(\varphi(\boldsymbol{p}))) = \left(\hat{R}_x(\boldsymbol{p})\hat{S}_y(\boldsymbol{p}) - \hat{R}_y(\boldsymbol{p})\hat{S}_x(\boldsymbol{p})\right)^2. \tag{5}$$

The overall distance between R and T is defined as

$$\begin{aligned} \mathcal{D}(R, T) &= \frac{1}{2} \int_\Omega d\left(R(\boldsymbol{p}), T(\varphi(\boldsymbol{p}))\right) d\boldsymbol{p} \\ &\approx \frac{h^2}{2} \sum_{\boldsymbol{p} \in \Omega} \left(\hat{R}_x(\boldsymbol{p})\hat{S}_y(\boldsymbol{p}) - \hat{R}_y(\boldsymbol{p})\hat{S}_x(\boldsymbol{p})\right)^2. \end{aligned} \tag{6}$$

Where h is the distance between grid points, defined as the image domain Ω, divided by the number of grid points in each direction.

A Gauss-Newton optimization scheme is used to solve for the registration parameters $\boldsymbol{\gamma}$. For each iteration the parameter change $\boldsymbol{s}$ is found that minimizes $f(\boldsymbol{\gamma} + \boldsymbol{s})$. Using a second order Taylor expansion we obtain

$$f(\boldsymbol{\gamma} + \boldsymbol{s}) \approx f(\boldsymbol{\gamma}) + \nabla f(\boldsymbol{\gamma}) \cdot \boldsymbol{s} + \frac{1}{2} \boldsymbol{s} \cdot \nabla^2 f(\boldsymbol{\gamma}) \cdot \boldsymbol{s}. \tag{7}$$

This equation needs to be minimized; therefore the first order derivative of the function with respect to $\boldsymbol{s}$ should equal zero.

$$f'(\boldsymbol{\gamma} + \boldsymbol{s}) \approx \nabla f(\boldsymbol{\gamma}) + \nabla^2 f(\boldsymbol{\gamma}) \cdot \boldsymbol{s}. \tag{8}$$

By setting the result equal to zero and rearranging the terms we obtain

$$-\nabla f(\boldsymbol{\gamma}) = \nabla^2 f(\boldsymbol{\gamma}) \cdot \boldsymbol{s}. \tag{9}$$

$f(\boldsymbol{\gamma})$ can be written in a general form:

$$f(\boldsymbol{\gamma}) = \frac{h^2}{2}|\boldsymbol{r}(\boldsymbol{\gamma})|_2^2, \quad \text{with} \quad \boldsymbol{r}(\boldsymbol{\gamma_p}) = \hat{R}_x(\boldsymbol{p})\hat{S}_y(\boldsymbol{p}) - \hat{R}_y(\boldsymbol{p})\hat{S}_x(\boldsymbol{p}). \tag{10}$$

The matrix of the first-order partial derivatives of $\boldsymbol{r}$ is the Jacobian,

$$J = [\frac{\partial \boldsymbol{r}_j}{\partial \boldsymbol{\gamma}_p}]. \tag{11}$$

By using the Jacobian the derivatives of $f(\boldsymbol{\gamma})$ can be written as:

$$\nabla f(\boldsymbol{\gamma}) = J(\boldsymbol{\gamma})^T \boldsymbol{r}(\boldsymbol{\gamma}). \tag{12}$$

$$\nabla^2 f(\boldsymbol{\gamma}) = J(\boldsymbol{\gamma})^T J(\boldsymbol{\gamma}) + \sum_{j=1}^{2} \boldsymbol{r}_j(\boldsymbol{\gamma}) \nabla^2 \boldsymbol{r}_j(\boldsymbol{\gamma}) \approx J(\boldsymbol{\gamma})^T J(\boldsymbol{\gamma}). \tag{13}$$

Substituting Eqn. 12 and 13 in Eqn. 9 and rearranging the terms leads to

$$\boldsymbol{s} = -\big(J(\boldsymbol{\gamma})^T J(\boldsymbol{\gamma})\big)^{-1} J(\boldsymbol{\gamma})^T \boldsymbol{r}(\boldsymbol{\gamma}). \tag{14}$$

The partial derivates $\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\gamma}}$ of $\boldsymbol{r}(T(\varphi(\boldsymbol{\gamma})))$ can be calculated using the chain rule

$$\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\gamma}} = \frac{\partial \boldsymbol{r}}{\partial T} \frac{\partial T}{\partial \varphi} \frac{\partial \varphi}{\partial \boldsymbol{\gamma}}, \tag{15}$$

leading to

$$\frac{\partial \boldsymbol{r}}{\partial \boldsymbol{\gamma}} = \begin{pmatrix} -\hat{R}_y & \hat{R}_x \end{pmatrix} \begin{pmatrix} \hat{S}_{xx} & \hat{S}_{xy} \\ \hat{S}_{xy} & \hat{S}_{yy} \end{pmatrix} \begin{pmatrix} -x_1 \sin(\gamma_1) - x_2 \cos(\gamma_1) & 1 & 0 \\ x_1 \cos(\gamma_1) - x_2 \sin(\gamma_1) & 0 & 1 \end{pmatrix}.$$

We use a convolution with a Gaussian kernel to calculate image derivatives.

$$\phi_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{1}{2}(x^2+y^2)/\sigma^2}. \tag{16}$$

Derivatives of the image can be calculated at any scale by

$$DI(x, y, \sigma) = (D\phi_\sigma * I)(x, y), \tag{17}$$

where D is any linear derivative operator with constant coefficients. By convolving the reference image R with the first order derivatives of the Gaussian kernel we obtain $\hat{R}_x$ and $\hat{R}_y$ and similarly by convolving the transformed template T with the second order derivatives of the Gaussian kernel we obtain $\hat{S}_{xx}, \hat{S}_{xy}$, and $\hat{S}_{yy}$.

By iterating Eqn. 14 and updating the parameters $\boldsymbol{\gamma}_{i+1} = \boldsymbol{\gamma}_i + \boldsymbol{s}$ the best match between the template and the reference image is established. The process is repeated until the distance $\mathcal{D}(R, T)$ divided by the number of pixels in the template T, between two consecutive iterations, is less than $1e-9$. This threshold is experimentally set. A change of an order of magnitude in this threshold will not significantly change the outcome of the algorithm. This threshold is the same for all sequences.

The obtained parameters $\boldsymbol{\gamma}$ can be used to track the cervical vertebra from frame i to frame $i+1$. This is done for each vertebra and for the skull.

3 Results

The results of our automatic tracking algorithm are presented to the user in an animation as shown in Fig. 3. The results were validated by comparison with manual vertebrae outline tracking by a trained clinician. The manual tracking was obtained by outlining all cervical vertebrae and the skull in a neutral frame. In the consecutive frames the outlines of the previous frame are displayed over the image. These outlines are then rotated and translated horizontally and vertically by the user until they visually match with the corresponding vertebrae in that frame. This is repeated until all the frames contain matching outlines. This manual analysis of an entire anteflexion-flexion movie takes a trained expert 5 to 6 hours.

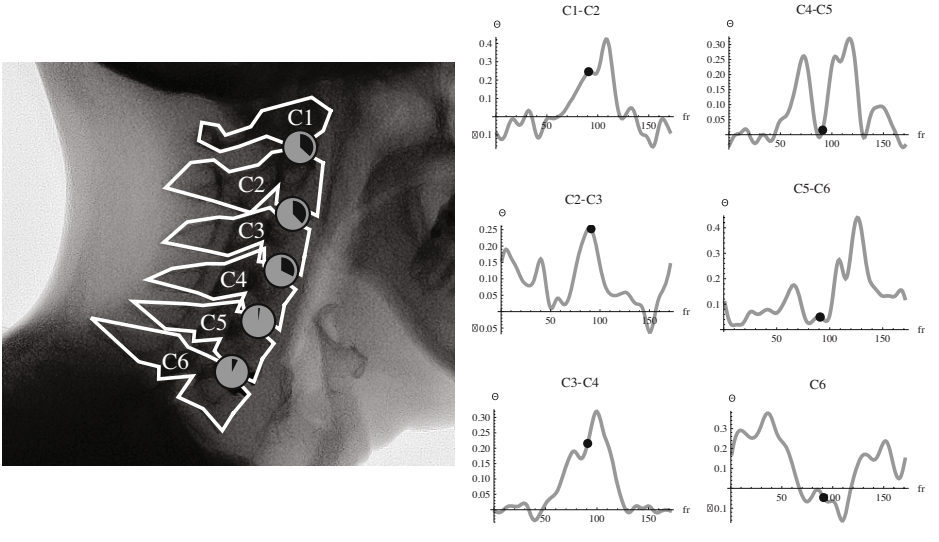


Fig. 3. Visualization of the results; the user is presented with an animation where on the left side the vertebrae and a moving outline (user-specified in a single frame) are shown; the amount of rotation between two adjacent vertebrae is visualized by (exaggerated) polar-wedge diagrams. On the right side the amount of rotation between adjacent vertebrae is quantified in graphs, the black dot represents the current frame.

4 Discussion

The angles of rotation obtained by both the expert and our algorithm were compared for three separate cervical fluoroscopy video's, of which one contains a disc prosthesis. In all cases the automatic method performed as well as, or better than the manual tracking by the expert. The performance of one sequence is shown in Fig. 4. It is clear that the differences are small apart from some manual annotation errors that are apparent as high peaks in the graphs.

Our algorithm is robust to variations in the choice of the templates (Sec. 2.2) as long as a substantial part of template is visible throughout the entire sequence. If, however, the template area is too small, a region without any structure or an area which will be largely occluded later in the video sequence is selected, the algorithm will fail to track that vertebra. As a test we have asked six different users to draw the templates for the vertebrae and the skull, and in all of these cases the algorithm was able to track the vertebrae flawlessly, even though there were substantial differences in the selected areas.

Our method is insensitive for the presence of a disc prosthesis, which is valuable as it enables the analysis of the movement of the cervical spine before and after the placement of such a prosthesis.

The frames of the video sequence were cropped such that only areas that contain information on the vertebrae and the skull were kept, i.e. the shoulders were removed from the image and any excess space in front of or above the patient was removed. On average this reduces the dimensions of the images to 500×500 pixels. Implementing these changes into the fluoroscopy acquisition protocol will reduce the dose.

Selecting the neutral frame and marking the template areas on the video sequence takes no more than 3 minutes and is all the user interaction required for the algorithm to run. Our *Mathematica*™ implementation of the automatic tracking algorithm takes around 45 minutes to run for all vertebrae on the entire fluoroscopy video sequence, this algorithm runs unsupervised.

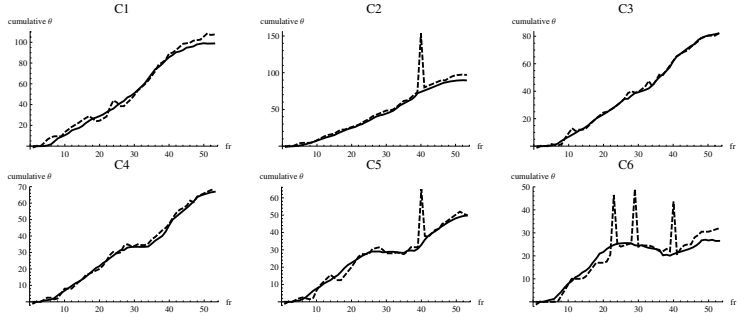


Fig. 4. The cumulative angle of each vertebra through the video sequence, as marked by the user (dashed) and our algorithm (continuous)

Processing time can be shortened by reducing the video frame rate. We were able to reduce the original frame rate of 10 frames per second to 3 frames per second without any loss of accuracy. (Note that adjusting the acquisition protocol to this frame rate will lead to a significant dose reduction). If the movement between two consecutive frames is too large, the algorithm will fail to track the vertebrae correctly. An algorithm to drop consecutive frames with little movement could further decrease the computation time.

The goal of this study was to develop an automated cervical vertebrae tracking algorithm that can give quantitative data on the order of movement of segments of the cervical spine. Our method is able to accurately track the skull and the cervical vertebrae throughout an entire cervical fluoroscopy video. It is less labor-intensive than manual methods, less complex and not as computationally demanding as existing methods which extract the contours of the vertebra in each frame of the video sequence. The tracking algorithm is robust and allows for large variations in the choice of the template regions.

References

1. van Mameren, H., Drukker, J., Sanches, H., Beursgens, J.: Cervical spine motion in the sagittal plane (i) range of motion of actually performed movements, an x-ray cinematographic study. Eur. J. Morphol. 28(1), 47–68 (1990)
2. van Mameren, H., Sanches, H., Beursgens, J., Drukker, J.: Cervical spine motion in the sagittal plane (ii). position of segmental averaged instantaneous centers of rotation–a cineradiographic study. Spine 17(5), 467–474 (1992)
3. Howe, B., Gururajan, A., Sari-Sarraf, H., Long, L.R.: Hierarchical segmentation of cervical and lumbar vertebrae using a customized generalized hough transform and extensions to active appearance models. In: Proc. 6th IEEE Southwest Symposium on Image Analysis and Interpretation, March 28–30, 2004, pp. 182–186 (2004)
4. Zheng, Y., Nixon, M., Allen, R.: Automated segmentation of lumbar vertebrae in digital videofluoroscopic images. IEEE Transactions on Medical Imaging 23(1), 45–52 (2004)
5. Roberts, M., Cootes, T., Adams, J.: Vertebral shape: Automatic measurement with dynamically sequenced active appearance models. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, p. 733. Springer, Heidelberg (2005)
6. Chaitanya, B.: Segmentation of radiographs of cervical spine using level sets. Master's thesis, Texas Tech University (2006)
7. Benjelloun, M., Mahmoudi, S.: X-ray image segmentation for vertebral mobility analysis. International Journal of Computer Assisted Radiology and Surgery 2(6), 371–383 (2008)
8. Haber, E., Modersitzki, J.: Intensity gradient based registration and fusion of multimodal images. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 726–733. Springer, Heidelberg (2006)

Expertise Modeling for Automated Planning of Acetabular Cup in Total Hip Arthroplasty Using Combined Bone and Implant Statistical Atlases

Itaru Otomaru[1], Kazuto Kobayashi[1], Toshiyuki Okada[2], Masahiko Nakamoto[3], Yoshiyuki Kagiyama[4], Masaki Takao[5], Nobuhiko Sugano[6], Yukio Tada[1], and Yoshinobu Sato[3]

[1] Graduate School of Engineering, Kobe University
[2] Medical Center for Translational Research, Osaka University Hospital
[3] Department of Radiology, Graduate School of Medicine, Osaka University
[4] The Center for Advanced Medical Engineering and Informatics, Osaka University
[5] Department of Orthopaedic Surgery, Graduate School of Medicine, Osaka University
[6] Department of Orthopaedic Medical Engineering, Graduate School of Medicine, Osaka University

Abstract. Intraoperative robotic and computer-guided assistances are now commonly used in total hip arthroplasty (THA) for accurate execution of the preoperative plan. Although the preoperative plan to be accurately executed is critical, it is still interactively prepared in a time-consuming and subjective manner. In this paper, atlas-based approach to automated surgical planning of the acetabular cup in THA is described to stabilize its quality as well as reduce its time-consuming nature. Surgeon's expertise is embedded in two types of statistical atlases, which are constructed from training datasets of CT-based 3D plans prepared by experienced surgeons. One is a statistical shape model which encodes global spatial relationships between the patient anatomy and implant. The other is the statistical map of residual bone thickness on the implant surface, which encodes local spatial constraints of the anatomy and implant. Given the 3D pelvis shape of the patient, we formulate a procedure to determine the best size and position of the acetabular cup which satisfy the constraints derived from the two statistical atlases. We validated the proposed planning method by retrospective study using the datasets which were actually used in the THA surgery.

1 Introduction

Surgical CAD/CAM is one of common frameworks for computer assisted surgery [1], especially, in orthopedic application. The CAM systems, that involve intraoperative robotic and computer-guided assistance, ensure accurate execution of preoperative plans constructed using the CAD systems. In its ultimate form, the surgery can be regarded as virtually completed once preoperative planning is completed. Therefore, the quality of preoperative planning is becoming more

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 532–539, 2009.

critical than ever in this framework [2] [3]. Our aim is to automate preoperative planning for total hip arthroplasty (THA) in order to stabilize its quality as well as reduce its time-consuming nature.

Nowadays, navigated THA is routinely used in many hospitals. Especially, for patients with osteoarthritis of the hip caused by congenital hip dysplasia, CT-based 3D preoperative planning is effective due to large deformation of the hip joint. A large number of the 3D preoperative plans prepared by well-experienced surgeons have been accumulated in some of the hospitals. The feedback of these past planning datasets can be utilized for the future planning. Closed-loop surgery to make the best use of past surgical data for future surgery is a recently emerged framework [4]. Atlas-based approach is well-fit to this framework and has been applied to the femoral stem planning [5] [6]. One of the simplest forms of such approach is to use one standard 3D plan constructed on CT data as a template [5]. The spatial relations between host bone (femur) and implant (femoral stem) in the template are mapped to each patient by scaled rigid registration between the patient and template. However, the template based on only one dataset is insufficient to deal with inter-patient shape variations. More recently, the statistical map of bone-implant distance on implant surface was generated from a large number of 3D plans to automate femoral stem planning [6].

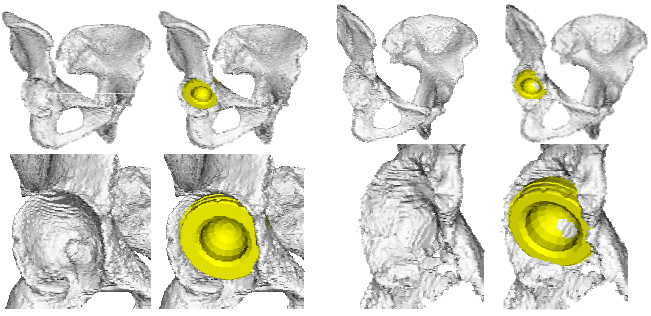
In this paper, we develop an atlas-based approach to automated acetabular cup planning. Unlike femoral stem planning, the expertise of cup planning involves more global relations between implant and patient anatomy. Therefore, we represent statistically-derived constraints of bone-implant relations not only as the form of a statistical map on implant surface but also as a statistical shape model [7] to describe global relations. We formulate a method for hybrid use of the two statistical atlases to automate cup planning, and evaluate it using planning datasets actually used in navigated THA in comparison with a previous automated cup planning method [8] based on user-specified constraints.

2 Methods

2.1 Overview

Figure 1 shows typical cup planning examples. The position and size of the acetabular cup should be basically determined so as to recover the original anatomy of the acetabulum. Although it is not so difficult to predict the original anatomy for a mildly diseased case (Fig. 1(a)), it is somewhat difficult for a severely diseased case due to its large deformation and shift (Fig. 1(b)). In both cases, cup size and position should be carefully refined so that residual bone thickness is sufficient on the implant cup surface and no penetration occurs.

Two types of statistical atlases are constructed from a sufficient number of cup plans on CT data prepared by experienced surgeons. One is a combined pelvis and cup statistical shape model (PC-SSM) to roughly predict the cup shape recovering the original anatomy based on global pelvic anatomy [9]. The other is the statistical map of (residual) bone thickness (SM-BT) on the cup surface for subsequent refinement of the cup size and position so as to ensure



(a) Mildly diseased case. (b) Severely diseased case.

Fig. 1. Cup planning for mildly and severely diseased pelvises. Cup planning was performed by an experienced surgeon. Left column: Pelvis model. Right column: Cup planning on pelvis model.

sufficient residual bone thickness for the host bone which was originally developed to describe the statistical map of bone-implant distance for femoral stem planning [6].

In our automated cup planning, given the pelvis surface model segmented from CT data, the cup position and size are estimated. The cup is modeled as a hemisphere and its diameter and center are regarded as the cup size and position, respectively. The final cup size is determined by selecting the size closest to the estimated diameter among available discrete cup sizes. Because cup orientation is determined based on femoral anteversion in our hospital, we do not address its determination in this paper where only the pelvis is considered.

2.2 Construction of Planning Atlases

We assume that N training datasets of cup planning are given. Let P_i and C_i be the 3D shape models of the pelvis and the cup model planned on P_i, respectively, where $i = 1, \cdots, N$. Inter-patient nonrigid registration of P_i and C_i is performed using a point based method. Let $\mathbf{p}_i$ and $\mathbf{c}_i$ be the sets of vertex's position of P_i and C_i, respectively. Then, we concatenate $\mathbf{p}_i$ and $\mathbf{c}_i$ to obtain the combined pelvis and cup pointset $\mathbf{q}_i$. The combined pelvis and cup statistical shape model (PC-SSM) is obtained by performing principal component analysis of $\mathbf{q}_i$. PC-SSM $\mathbf{q}$ is defined as

$$\mathbf{q}(\mathbf{b}) = \bar{\mathbf{q}} + \Phi\mathbf{b}, \quad \bar{\mathbf{q}} = \begin{bmatrix} \bar{\mathbf{p}} \\ \bar{\mathbf{c}} \end{bmatrix}, \tag{1}$$

where $\bar{\mathbf{q}}$, Φ, and $\mathbf{b}$ are the mean shape vector, the eigenvector matrix, and the shape parameter vector, respectively. $\bar{\mathbf{q}}$ consists of the mean shape vectors of the pelvis and the cup, $\bar{\mathbf{p}}$ and $\bar{\mathbf{c}}$, respectively.

The statistical map of bone thickness (SM-BT) is defined as the distributions of the mean and standard deviation of the residual bone thickness on the cup

surface. The residual bone thickness is defined as the remaining thickness of bone tissue when the bone is removed for cup placement. The residual bone thickness is calculated on each cup vertex along each normal direction. If the bone thickness is equal to or smaller than zero, it means that penetration of the cup through the bone occurs. Since the cup surface is not fully covered with the pelvic bone, the target area of SM-BT is defined as the area which are covered with the pelvic bone in 95% cases of the training datasets. Let $\mathbf{v}_{ij}$ be the j-th vertex's position of the cup surface of the i-th training dataset. Let $d(\mathbf{v}_{ij}, P_i)$ be the bone thickness at $\mathbf{v}_{ij}$. If the point $\mathbf{v}_{ij}$ is not covered with the pelvic bone P_i, the bone thickness is not calculated. The distributions of the mean and standard deviation of the bone thickness on the cup surface, U and S, are obtained by the method described in [3] and defined as

$$U = \{\mu_j | \mu_j = \frac{1}{N} \sum_{i=1}^{N} d(\mathbf{v}_{ij}, P_i), \quad \mathbf{v}_{ij} \in C_i\}, \tag{2}$$

$$S = \{\sigma_j | \sigma_j = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (d(\mathbf{v}_{ij}, P_i) - \mu_i)^2}, \quad \mathbf{v}_{ij} \in C_i\}, \tag{3}$$

where μ_j and σ_j represent the mean and standard deviation of the bone thickness at the point $\mathbf{v}_j$, respectively.

2.3 Optimization Procedure

Two-step optimization is performed for automated cup planning. Let P_a be the patient's 3D pelvis surface model on which the cup is placed. The cup position $\mathbf{t}$ and size s are initially determined by fitting the pelvis part of PC-SSM to P_a, and then they are refined by using SM-BT. When specific parameter $\mathbf{b}$ is given to PC-SSM $\mathbf{q}(\mathbf{b})$ in Eq. (1), the pelvis shape $\mathbf{p}(\mathbf{b})$ and the cup shape $\mathbf{c}(\mathbf{b})$ are given by $\bar{\mathbf{p}} + \Phi_{\mathbf{p}}\mathbf{b}$ and $\bar{\mathbf{c}} + \Phi_{\mathbf{c}}\mathbf{b}$, respectively, where $\Phi_{\mathbf{p}}$ and $\Phi_{\mathbf{c}}$ are the submatrices of Φ which correspond to pelvis and cup vertices, respectively. Let the number of vertices of $\mathbf{p}(\mathbf{b})$ and $\mathbf{c}(\mathbf{b})$ be n and m, respectively. The fitting process of PC-SSM is performed by obtaining $\mathbf{b}$ which minimizes the cost function C_D defined as

$$C_D(\mathbf{b}; \mathbf{p}(\mathbf{b}), P_a) = \sum_{j=1}^{n} d(\mathbf{w}_j(\mathbf{b}), P_a)^2, \tag{4}$$

where $\mathbf{w}_j(\mathbf{b})$ and $d(\mathbf{w}_j(\mathbf{b}), P_a)$ are the j-th vertex's position of $\mathbf{p}(\mathbf{b})$ and the shortest distance from $\mathbf{w}_j(\mathbf{b})$ to P_a, respectively. Let $\mathbf{b}_a$ be the shape parameter vector minimizing Eq. (4). We use Levenberg-Marquardt method for the minimization. Then, initial cup position $\mathbf{t}_a$ and size s_a are determined by fitting a parametric hemisphere model to $\mathbf{c}(\mathbf{b}_a)$. Let $\mathbf{c}'(\mathbf{t}, s)$ be the sets of vertex's position of the parametric hemisphere model which position is $\mathbf{t}$ and size is s.

In the subsequent refinement using SM-BT, we minimize the cost function $f(\mathbf{t}, s)$ defined as

$$f(\mathbf{t}, s) = \lambda_1(s - s_a)^2 + \lambda_2(\mathbf{t} - \mathbf{t}_a)^2 + \lambda_3 \frac{1}{m} \sum_{j=1}^{m} \frac{1}{\sigma_j^2} (d(\mathbf{v}'_j(\mathbf{t}, s), P_a) - \mu_i)^2, \quad (5)$$

where $\mathbf{v}'_j(\mathbf{t}, s)$ is the j-th vertex's position of $\mathbf{c}'(\mathbf{t}, s)$. The first and second terms ensure that the size and position estimated using PC-SSM should not be far from the initial size and position. The third term ensures that the distribution of the bone thickness on the ongoing cup becomes close to SM-BT. λ_1, λ_2, and λ_3 are the weight parameters balancing the three terms. For the hybrid use of the two atlases, $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, $\lambda_3 = 10.0$ were used while $\lambda_1 = 1.0$, $\lambda_2 = 1.0$, $\lambda_3 = 0.0$ or $\lambda_1 = 0.0$, $\lambda_2 = 0.0$, $\lambda_3 = 1.0$ were used when either atlas was utilized.

3 Experimental Results

3.1 Experimental Conditions

We compared the proposed method with the previous method described in [8], which is the only one existing method for automated cup planning in 3D to our knowledge. In order to investigate effects of the atlases, we tested the method under three conditions, that is, only PC-SSM, only SM-BT, and the hybrid use of PC-SSM and SM-BT. CT slice thickness and reconstruction pitch were 2 mm. Field of view was 360 mm. The pelvis surface was segmented with the atlas-based automated segmentation method described in [10]. The available cup size variation was 11 sizes from 40 mm to 60 mm. The 3D surface models of pelvis and cups consisted of 3,000 and 353 vertices, respectively. 28 patient datasets including 3D pelvis models and preoperative plans were used for construction of atlases and evaluations. The preoperative plans used for atlas construction and evaluation were made by the experienced surgeons using a commercially available planning system, Stryker CT-Hip System (Stryker Leibinger GmbH, Freiburg, Germany), and applied to the actual navigated THA surgery. We used the first 10 principal components in PC-SSM fitting to avoid over-fitting [9].

Evaluation was performed by leave-one-out cross validation. Evaluation measures were cup position error, cup size error and the number of cases when the penetration occurred. Error was defined as the difference between the surgeon's and estimated results.

3.2 Results

The constructed PC-SSM and SM-BT are shown in Fig. 2 and Fig. 3, respectively. Fig. 2 shows the mean and variations of PC-SSM. Fig. 3(a) shows the evaluation area for SM-BT construction. Fig. 3(b) and (c) show the distribution of the mean and standard deviation of the bone thickness, respectively. Table. 1 shows a summary of the evaluation results. It shows the number of cases when the penetration of the cup occurred, mean positional error, and mean size error.

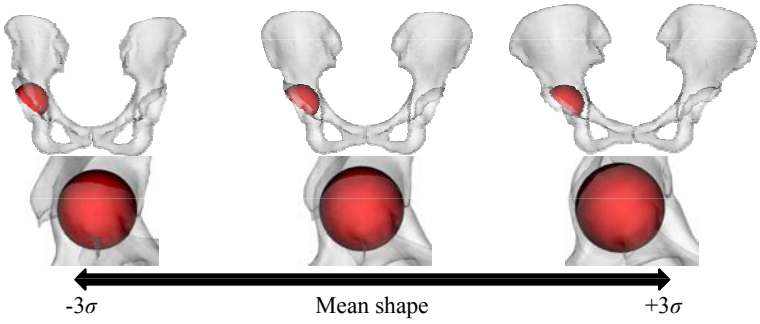


Fig. 2. Shape variations of combined pelvis and cup statistical shape model (PC-SSM). The first mode is shown.

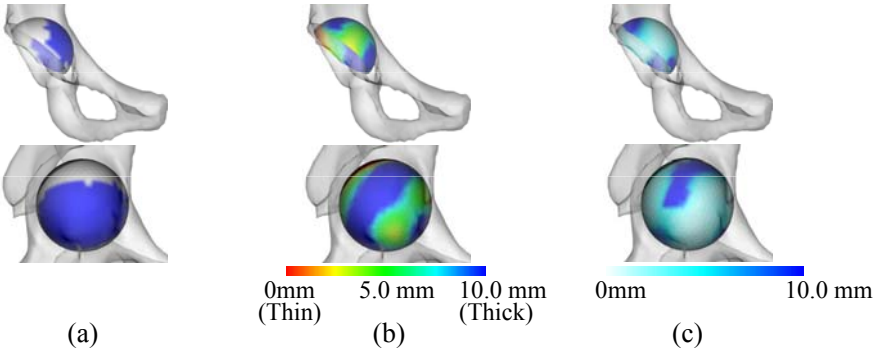


Fig. 3. Statistical map of residual bone thickness (SM-BT). (a) Blue area indicates the cup coverage over 95 % of datasets. (b) Distribution of mean bone thickness. (c) Distribution of standard deviation of bone thickness. Color bars at bottom of each figure represent the residual bone thickness.

With the previous method, there was no cup penetration, but the mean positional error was the largest among the four conditions, which was 4.3 mm. With the PC-SSM only, the size error was 1.5 mm, which was smaller than the previous method. On the other hand, cup penetration occurred in four cases. With the SM-BT only, the number of cases of penetration was reduced, but the size error was 2.4 mm which was the largest among the four conditions. With the hybrid use of PC-SSM and SM-BT, both positional and size errors were the smallest among the four conditions, which were 3.9 mm and 1.4 mm, respectively. Additionally, only one penetration occurred.

Figure 4 shows the planning results of a typical case. With the PC-SSM only, the cup size was same as the surgeon's selection, but there was cup penetration. With the SM-BT only, positional error was the smallest, but larger size was selected. Additionally, there was penetration. On the other hand, with the hybrid use of PC-SSM and SM-BT, the positional and size error was smaller than that of PC-SSM only. Further, there was no penetration.

Table 1. Evaluation results of automated planning

	Previous	PC-SSM only	SM-BT only	Hybrid use of PC-SSM and SM-BT
Number of cup penetration	0	4	1	1
Mean positional error [mm]	4.3	4.2	4.0	3.9
Mean size error [mm]	2.1	1.5	2.4	1.4

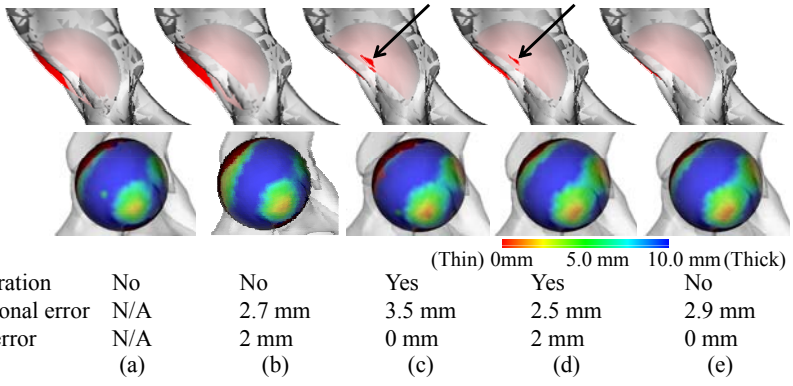


Fig. 4. Illustrative case of experimental results. (a) Surgeon's plan. (b) Previous method. (c) PC-SSM only. (d) SM-BT only. (e) Hybrid use of PC-SSM and SM-BT. Values below each figure indicate penetration occurred or not (top), positional error [mm] (middle), and size error [mm] (bottom), respectively. Arrows indicate the areas where penetration occurred. Colors on cup surface indicate bone thickness distribution.

4 Discussion and Conclusions

We proposed a method for hybrid use of the two atlases to automate 3D planning for acetabular cup placement in THA. One is a combined pelvis and cup statistical shape model (PC-SSM) which encodes global spatial relationships between the patient anatomy and implant. The other is a statistical map of residual bone thickness (SM-BT) which encodes local spatial constraints of the anatomy and implant. The proposed method could provide a framework for learning and modeling the planning policy of the surgeon because the two atlases are constructed from training datasets of experienced surgeons. It involves few manual parameter tuning while the previous method has many parameters. According to the experimental results, mean errors in size and position were better for the hybrid use of PC-SSM and SM-BT than the previous method (Table. 1). It would be arguable that this improvements is clinically significant or not. Currently, we have already had some preliminary results to evaluate clinical significance, that is, comparison with inter-surgeon difference. The mean inter-surgeon difference for eight cases was 1.3 mm in size while the mean differences from a surgeon of the proposed and the previous methods were 1.5 mm and 2.0 mm,

respectively. These results may show some implication on clinical significance. We will evaluate the statistical significance of these results as future work.

In the previous method, there was no penetration since it explicitly prohibited the penetration. However, cup penetration occurred in four cases when only PC-SSM was used. We consider this is because the variation of PC-SSM described by the training datasets is insufficient for fitting of severely diseased cases. On the other hand, with the hybrid use of PC-SSM and SM-BT, the number of cases of penetration was reduced. We consider that this is because the cup size and position were determined so as to minimize the difference between the resulting bone thickness and SM-BT which had no penetration. However, the current atlas could not prohibit all of the penetrations completely. The constraint which explicitly prohibits penetrations could be introduced to overcome the problem.

In principle, given a sufficient number of planning datasets that a surgeon planned, the method is applicable to various implants for different bones. As future work, we are planning to apply the method to the femoral stem.

References

1. Taylor, R.H., Stoianovici, D.: Medical Robotics in Computer-Integrated Surgery. IEEE T. Robot. Aut. 19(5), 765–781 (2003)
2. Bargar, W.L., Bauer, A., Borner, M.: Primary and Revision Total Hip Replacement using the Robodoc System. Clin. Orthop. Relat. R 354, 82–91 (1998)
3. Lattanzi, R., Viceconti, M., Zannoni, C., Quadrani, P., Toni, A.: Hip-Op: An Innovative Software to Plan Total Hip Replacement Surgery. Med. Inform. Internet. 27(2), 71–83 (2002)
4. Taylor, R.H., Menciassi, A., Fichtinger, G., Dario, P.: Medical Robotics and Computer-Integrated Surgery. Springer Handbook of Robotics, 1199–1222 (2008)
5. Viceconti, M., Testi, D., Simeoni, M., Zannoni, C.: An Automated Method to Position Prosthetic Components within Multiple Anatomical Spaces. Comput. Meth. Prog. Bio. 70(2), 121–127 (2003)
6. Nakamoto, M., Otomaru, I., Takao, M., Sugano, N., Kagiyama, Y., Yoshikawa, H., Tada, Y., Sato, Y.: Construction of Statistical Surgical Planning Model for Automated 3D Planning of Femoral Component in Total Hip Arthroplasty (THA). In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 718–725. Springer, Heidelberg (2008)
7. Cootes, T.F., Hill, A., Taylor, C.J., Haslam, J.: The Use of Active Shape Models for Locating Structures in Medical Images. In: Barrett, H.H., Gmitro, A.F. (eds.) IPMI 1993. LNCS, vol. 687, pp. 33–47. Springer, Heidelberg (1993)
8. Kagiyama, Y., Nakamoto, M., Takao, M., Sato, Y., Sugano, N., Yoshikawa, H., Tamura, S.: Automated Preoperative 3D Planning of Acetabular Cup Positioning and Size Selection in Total Hip Arthroplasty using CT Data. In: Proc. of CAOS 2004, pp. 312–313 (2004)
9. Otomaru, I., Kobayashi, K., Okada, T., Nakamoto, M., Takao, M., Sugano, N., Tada, Y., Sato, Y.: CT-based Automated Preoperative Planning of Acetabular Cup Size and Position using Pelvis-cup Integrated Statistical Shape Model. In: Proc. of CAOS 2009, pp. 185–188 (2009)
10. Okada, T., Shimada, R., Hori, M., Nakamoto, M., Chen, Y., Nakamura, H., Sato, Y.: Automated Segmentation of the Liver from 3D CT Images Using Probabilistic Atlas and Multilevel Statistical Shape Model. Acad. Radiol. 15(1), 1390–1403 (2008)

Wide-Angle Intraocular Imaging and Localization*

Christos Bergeles[1], Kamran Shamaei[1], Jake J. Abbott[1,2], and Bradley J. Nelson[1]

[1] Institute of Robotics and Intelligent Systems, ETH Zurich, 8092 Zurich, Switzerland
{cbergeles,kamrans,bnelson}@ethz.ch
[2] Department of Mechanical Engineering, University of Utah, Salt Lake City, Utah 84112
jake.abbott@utah.edu

Abstract. Vitreoretinal surgeries require accuracy and dexterity that is often beyond the capabilities of human surgeons. Untethered robotic devices that can achieve the desired precision have been proposed, and localization information is required for their control. Since the interior of the human eye is externally observable, vision can be used for their localization. In this paper we examine the effects of the human eye optics on imaging and localizing intraocular devices. We propose a method for wide-angle intraocular imaging and localization. We demonstrate accurate localization with experiments in a model eye.

1 Introduction

Future ophthalmic surgeries will be partially automated in order to achieve the accuracy needed for operations near the sensitive structure of the human retina. Our work is motivated by the untethered intraocular microrobot presented in [1]. The microrobot can move in the posterior of the human eye and is magnetically controlled; for accurate control, knowledge of its position inside the magnetic field is necessary [2]. Since the interior of the human eye is externally observable, vision-based 3D localization can be performed.

Intraocular imaging has been practiced for centuries, and clinicians now have the ability to acquire high-definition, magnified images of the interior of the eye using optical tools that are designed specifically for the human eye. However, these tools combined with the optical elements of the eye (Fig. 1(a)) alter the formation of images (Fig. 1(b)) and make localization challenging.

Our approach to intraocular localization is based on depth-from-focus [3]. Focus-based methods do not require a model of the object of interest, but only knowledge of the optical system. Focus-based methods applied in the eye could

* This work was supported by the NCCR Co-Me of the Swiss National Science Foundation. Part of this work was carried out while J. J. Abbott was with the Institute of Robotics and Intelligent Systems.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 540–548, 2009.

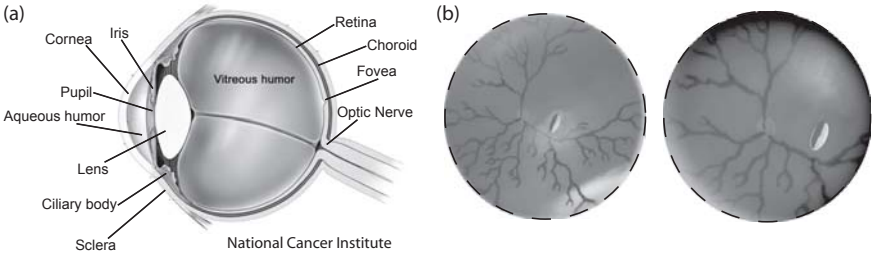


Fig. 1. (a) Anatomy of the human eye. (b) The biomedical microrobot of [1] in the model eye [4]. The left image shows the intraocular environment without the eye's optical elements, and the right image shows the effect of the model eye optics. Images are taken with an unmodified digital camera.

also localize unknown objects such as floaters. As a result, our analysis need not be considered only in the scope of microrobot localization, but is applicable on any type of unknown foreign bodies.

In the following section, we consider a stationary camera and evaluate different ophthalmoscopic techniques based on their imaging and localizing capabilities. In Sec. 3 we present a wide-angle localization algorithm. Our experimental results can be found in Sec. 4, and we conclude with an outlook in Sec. 5.

2 Comparison of Ophthalmoscopy Methods

Our results are based on Navarro's schematic eye [5] (i.e. an optical model based on biometric data that explains the optical properties of the human eye). Navarro's schematic eye performs well for angles up to 70° measured from the center of the pupil and around the optical axis. For greater angles, the biometric data of each patient should be considered individually. Simulations are carried out with the OSLO optical lens design software. Throughout this section, the object's depth z is measured along the optical axis. We begin by investigating the feasibility of imaging and localizing intraocular devices using existing ophthalmoscopy methods.

2.1 Direct Ophthalmoscopy

In a relaxed state, the retina is projected through the eye optics as a virtual image at infinity. An imaging system can capture the parallel beams to create an image of the retina. In direct ophthalmoscopy the rays are brought in focus on the observer's retina [6]. By manipulating the formulas of [7] the field-of-view for direct ophthalmoscopy is found as 10° (Fig. 2(a)).

Every object inside the eye creates a virtual image. These images approach infinity rapidly as the object approaches the retina. Figure 3 (solid line) displays the distance where the virtual image is formed versus different positions of an intraocular object. In order to capture the virtual images that are created from objects close to the retina, an imaging system with near to infinite working

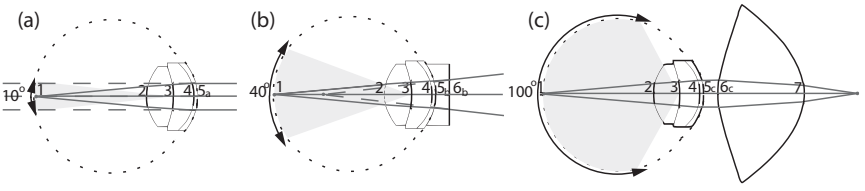


Fig. 2. (a) Direct ophthalmoscopy with Navarro's schematic eye [5]. (b) Ophthalmoscopy with Navarro's schematic eye with a vitrectomy lens [8]. (c) Indirect ophthalmoscopy with Navarro's schematic eye with a condensing lens [9].

Table 1. Optical parameters for the systems of Fig. 2

Surface	1	2	3	4	5a	5b	5c	6b	6c	7
Radius (mm)	12.00	6.00	−10.20	−6.50	−7.72	−7.72	−7.72	∞	11.65	−9.48
Conic Constant	0.00	−1.00	−3.13	0.00	−0.26	−0.26	−0.26	0.00	−9.24	−1.07
Thickness (mm)	16.32	4.00	3.05	0.55	∞	2.00	2.00	∞	13.00	∞
Refraction Index	1.336	1.420	1.337	1.376	1.000	1.425	1.000	1.000	1.523	1.000

distance is required. Such an imaging system will also have a large depth-of-field, and depth information from focus would be insensitive to object position.

2.2 Vitrectomy Lenses

To visualize devices operating in the vitreous humor of phakic (i.e. intact intraocular lens) eyes, only plano-concave lenses (Fig. 2(b)) need to be considered [6]. Vitrectomy lenses cause the virtual images of intraocular objects to form inside the eye, allowing the imaging systems to have a reduced working distance. Based on data given from HUCO Vision SA for the vitrectomy lens S5.7010 [8], we simulated the effects of a plano-concave vitrectomy lens on Navarro's eye (Fig. 2(b)). This lens allows for a field-of-view of 40°, significantly larger than the one obtainable with the method described in Sec. 2.1.

As shown in Fig. 3 (dashed line), the virtual images are formed inside the eye and span a lesser distance. Thus, contrary to direct observation, imaging with an optical microscope (relatively short working distance and depth-of-field) is possible. The working distance of such a system must be at least 20 mm. As depth-of-field is proportional to working distance, there is a fundamental limit to the depth-from-focus resolution achievable with vitrectomy lenses.

2.3 Indirect Ophthalmoscopy

Indirect ophthalmoscopy (Fig. 2(c)) allows for a wider field of the retina to be observed. A condensing lens is placed in front of the patient's eye, and catches rays emanating from a large retinal area. These rays are focused after the lens, creating an aerial image of the patient's retina. Condensing lenses compensate for the refractive effects of the eye, and create focused retinal images.

We simulated the effects of a double aspheric condensing lens based on information found in [9]. This lens, when placed 5 mm from the pupil, allows

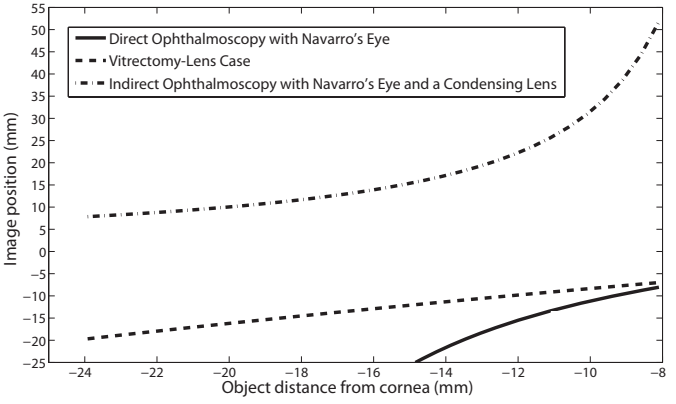


Fig. 3. Image position versus intraocular object position for the direct ophthalmoscopy case, the vitrectomy-lens case, and the indirect ophthalmoscopy case. Image distances are measured from the final surface of each optical system (5a, 6b, 7 respectively).

imaging of the peripheral retina and offers a field-of-view of 100°. As a result, it can be part of an imaging system with a superior field-of-view than the ones described in Sec. 2.1 and Sec. 2.2. The image positions versus the intraocular object positions can be seen in Fig. 3 (dashed-dotted line). A sensing system with a short working-distance and shallow depth-of-field can be used in order to extract depth information from focus for all areas inside the human eye. Depth estimation is more sensitive for objects near the intraocular lens, since smaller object displacements result in larger required focusing motions.

Dense CMOS sensors have a shallow depth-of-focus, and as a result, they can be used effectively in depth-from-focus techniques. Based on Fig. 3, to localize objects in the posterior of the eye a sensor travel of 10 mm is necessary. A $24 \times 24\,\text{mm}^2$ CMOS sensor can capture the full field-of-view. The simulated condensing lens causes a magnification of $0.78\times$ and thus, a structure of $100\,\mu$m on or near the retina will create an image of $78\,\mu$m. Even with no additional magnification, a CMOS sensor with a common sensing element size of $6 \times 6\ \mu\text{m}^2$ will resolve small retinal structures sufficiently. As a conclusion, direct sensing of the aerial image leads to a high field-of-view, while having advantages in focus-based localization.

3 Wide-Angle Localization

As previously stated, the condensing lens projects the spherical surface of the retina onto a flat aerial image. Moving the sensor with respect to the condensing lens focuses the image at different surfaces inside the eye, which we call *isofocus* surfaces. The locus of intraocular points that are imaged on a single pixel is called an *isopixel* curve. Figure 4(a) shows a subset of these surfaces and curves and their fits for the system of Fig. 2(c). The position of an intraocular point is found as the intersection of its corresponding isopixel curve and isofocus surface.

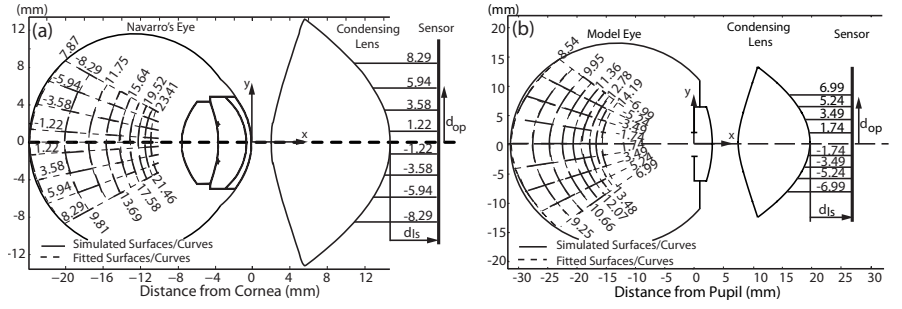


Fig. 4. Simulation of the isofocus surfaces and isopixel curves for (a) indirect ophthalmoscopy with Navarro's eye, and (b) indirect ophthalmoscopy with the model eye [4]. The different isofocus surfaces correspond to the distance from the lens to the sensor (d_{ls}), for uniform sensor steps of (a) ~ 1.95 mm, and (b) ~ 0.7 mm. The isopixel curves correspond to pixel distances from the optical axis (d_{op}), for uniform steps of (a) ~ 2.25 mm, and (b) ~ 1.75 mm.

The location of the isofocus surfaces and isopixel curves are dependent on the condensing lens and the individual eye. The optical elements of the human eye can be biometrically measured. For example, specular reflection techniques or interferometric methods can be used to measure the cornea [10], and autokeratometry or ultrasonometry can be used to measure the intraocular lens [11]. Then, the surfaces and curves can be accurately computed offline using raytracing. In theory there is an infinite number of isofocus surfaces and isopixel curves, but in practice there will be a finite number due to the resolution of sensor movement and pixel size, respectively.

The density of the isofocus surfaces for uniform sensor steps in Fig. 4(a) demonstrates that the expected depth resolution is higher for regions far from the retina. The isopixel curves show that the formed image is inverted, and from their slope it is deduced that the magnification of an intraocular object increases farther from the retina. As a result, we conclude that both spatial and lateral resolutions increase for positions farther from the retina.

The isofocus surfaces result from the optics of a rotationally symmetric and aligned system composed of conic surfaces. We therefore assume that they are conic surfaces as well, which can be parametrized by their conic constant, curvature, and intersection with the optical axis. Since the isofocus surfaces correspond to a specific sensor position, their three parameters can also be expressed as functions of the sensor position.

The isopixel curves are lines, and it is straightforward to parametrize them using their slope and their distance from the optical axis at the pupil. Each isopixel curve corresponds to one pixel on the image, and its parameters are functions of the pixel's offset (measured from the image center) due to the rotational symmetry of the system. For the 2D case, two parameters are required.

In Fig. 5(a)-(d) the parametrizing functions of the isofocus surfaces and isopixel curves are displayed. The conic constant need not vary (fixed at -0.5) because it was observed that the surface variation can be successfully captured

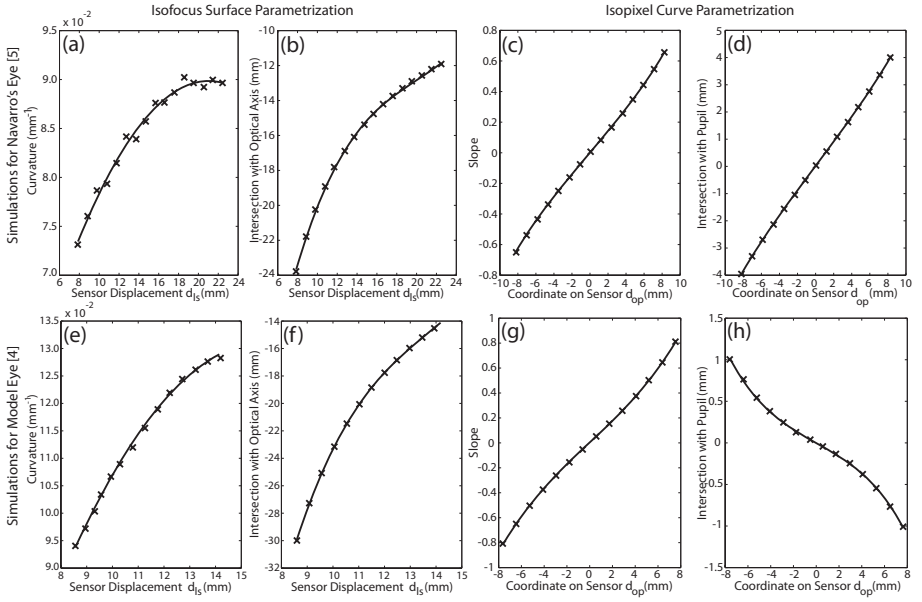


Fig. 5. (top row) Parametrization polynomials for the system of Fig. 4(a), and (bottom row) for the system of Fig. 4(b). Isofocus surface parametrization: (a), (b), (e), (f) Fitted 2nd- and 3rd-order polynomials for the curvature and for the intersection with the optical axis. Isopixel curve parametrization: (c), (d), (g), (h) Fitted 3rd-order polynomials for the line slope and for the intersection with the pupil.

by the curvature. For each parameter, we fit the least-order polynomial that captures its variability. The parametrizing functions are "$1-1$" (Fig. 5(a)-(d)), and thus, 3D intraocular localization with a wide-angle is unambiguous.

4 Localization Experiments

As an experimental testbed, we use the model eye [4] from Gwb International, Ltd. This eye is equipped with a plano-convex lens of 36 mm focal length that mimics the compound optical system of the human eye. Gwb International, Ltd. disclosed the lens' parameters so that we can perform our simulations. We also measured the model's retinal depth and shape.

The optical system under examination is composed of this model eye and the condensing lens of Fig. 2(c), where the refraction index was chosen as 1.531. The simulated isofocus surfaces and isopixel curves of the composite system, together with their fits, are shown in Fig. 4(b). Based on these simulations, we parametrize the isofocus surfaces and the isopixel curves (Fig. 5(e)-(h)). The behavior of the parameters is similar to the one displayed in Fig. 5(a)-(d) for Navarro's schematic eye. We assume an invariant conic constant of -1.05, because the variability of the surfaces can be captured sufficiently by the curvature.

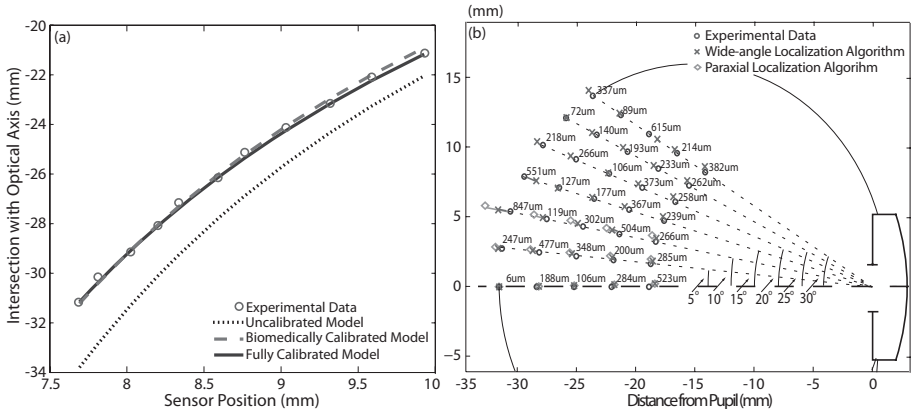


Fig. 6. (a) Model fits for the function describing the intersection of the isofocus surfaces with the optical axis. Biometric calibration errors: mean = 159 μm, std = 94 μm. (b) Localization experiment showing the validity of the proposed algorithm. Errors: mean = 282 μm, std = 173 μm.

In order to calibrate the isofocus surfaces for their intersection with the optical axis, we perform an on-optical-axis depth-from-focus experiment on the aligned optical system. We use a Sutter linear micromanipulation stage to move a checkerboard calibration pattern in the model eye with 1 mm steps, and estimate the in-focus sensor position [12]. The estimated sensor positions with respect to different object depths can be seen in Fig. 6(a). The uncalibrated model fit is displayed with a solid line, and, as can be seen, calibration is needed.

In the model eye, we can calibrate for the relationship between the in-focus sensor position and the depth of the object using the full set of data points. However, such an approach would be clinically invasive as it would require a vitrectomy and a moving device inside the eye. The only minimally invasive biometric data available are the depth and shape of the retina that can be measured from MRI scans [13]. Assuming that there are accumulated errors that can be lumped and included as errors in the estimated image and object positions, it is shown in [14] that by using a first-order model of the optics, calibration using only the depth of the retina is possible. By adapting this method to our framework, we are able to biometrically calibrate for the parameters of the polynomial that describes the intersection of the isofocus surfaces with the optical axis. The resulting fit can be seen in Fig. 6(a).

The remaining two parameters of the isofocus surfaces control the shape of the isofocus surfaces but not their position. The condensing lens is designed to create a flat aerial image of the retinal surface, and our experiments have shown that we can use it to capture an overall sharp image of the model eye's retina. Therefore, we conclude that there exists an isofocus surface that corresponds to the retinal surface, and we consider it as the 1$^{\text{st}}$ surface. From Fig. 4(b) we see that the 1$^{\text{st}}$ isofocus surface does indeed roughly correspond to the retinal shape (mean error = 371 μm). As a result, calibration for the conic constant and the

curvature is not needed. If our model was not accurately predicting the shape of the retina, then we would calibrate the parameters of the 1st isofocus surface so that is has exactly the same shape as the retina.

To estimate the validity of the presented wide-angle localization algorithm, we consider points in the model eye for various angles with respect to the optical axis and various distances from the pupil. In Fig. 6(b), the results using the proposed wide-angle localization algorithm are displayed. For comparison, we also show the results based on the paraxial localization algorithm presented in [14] for angles up to 10° from the optical axis. The paraxial localization results deteriorate as the angles increase. However, the localization method proposed here can be used for regions away from the optical axis with high accuracy.

5 Conclusions and Future Work

In this paper we evaluated existing ophthalmoscopy techniques based on their imaging and localizing capabilities. We presented the first wide-angle intraocular localization algorithm, and demonstrated accurate localization in a model eye (error: $282 \pm 173\,\mu$m). We are currently interested in localizing microrobots ($\sim 500\,\mu$m) for use in an open-loop magnetic field controller [1], and such localization errors are acceptable. Then, closed-loop visual servoing techniques will be used.

In the future, we will evaluate the robustness of the presented approach with respect to variations in illumination and uncertainties in the optical system. This is an important step towards *in vivo* intraocular localization.

References

1. Yesin, K.B., Vollmers, K., Nelson, B.J.: Modeling and control of untethered biomicrorobots in a fluidic environment using electromagnetic fields. Int. J. Robot. Res. 25(5-6), 527–536 (2006)
2. Abbott, J.J., Ergeneman, O., Kummer, M.P., Hirt, A.M., Nelson, B.J.: Modeling magnetic torque and force for controlled manipulation of soft-magnetic bodies. IEEE Trans. Robot. 23(6), 1247–1252 (2007)
3. Ens, J., Lawrence, P.: An investigation of methods for determining depth from focus. IEEE Trans. Pattern Anal. Machine Intell. 15(2), 97–108 (1993)
4. Gwb International, Ltd.: Model eye (2mm pupil) (2009), http://www.gwbinternational.com/model_eye.htm
5. Escudero-Sanz, I., Navarro, R.: Off-axis aberrations of a wide-angle schematic eye model. J. Opt. Soc. A. 16(8), 1881–1891 (1999)
6. Snead, M.P., Rubinstein, M.P., Jacobs, P.M.: The optics of fundus examination. Sur. Ophthalm. 36(6), 439–445 (1992)
7. Smith, G., Atchison, D.A.: The eye and visual optical instruments. Cambridge University Press, Cambridge (1997)
8. FCI Ophthalmics: S5.7010 planoconcave lens (2009), http://www.fci-ophthalmics.com/html/retina.html#lenses
9. Volk, D.A.: Indirect ophthalmoscopy lens for use with split lamp or other biomicroscope (January 6, 1998) U.S. Patent 5,706,073

10. Mejia-Barbosa, Y., Malacara-Hernandez, D.: A review of methods for measuring corneal topography. J. Opt. Vis. Sci. 78(4), 240–253 (2001)
11. Kirschkamp, T., Dunne, M., Barry, J.C.: Phakometric measurement of ocular surface radii of curvature, axial separations and alignment in relaxed and accommodated human eyes. Ophthal. Phys. Opt. 24(2), 65 (2004)
12. Sun, Y., Duthaler, S., Nelson, B.J.: Autofocusing in computer microscopy: selecting the optimal focus algorithm. J. Microsc. Res. Tech. 65(3), 139–149 (2004)
13. Atchison, D.A., Pritchard, N., Schmid, K.L., Scott, D.H., Jones, C.E., Pope, J.M.: Shape of the retinal surface in emmetropia and myopia. Invest. Ophthalm. & Vis. Sci. 46(8), 2698–2707 (2005)
14. Bergeles, C., Shamaei, K., Abbott, J.J., Nelson, B.J.: On imaging and localizing untethered intraocular devices with a stationary camera. In: IEEE Int. Conf. Biomed. Robot. Biomech., pp. 489–494 (2008)

Inverse C-arm Positioning for Interventional Procedures Using Real-Time Body Part Detection*

Christian Schaller, Christopher Rohkohl, Jochen Penne, Michael Stürmer, and Joachim Hornegger

Chair of Pattern Recognition and Erlangen Graduate School in Advanced Optical Technologies (SAOT), Friedrich-Alexander University Erlangen-Nuremberg, Martensstr. 3, 91058 Erlangen, Germany
christian.schaller@informatik.uni-erlangen.de

Abstract. The automation and speedup of interventional therapy and diagnostic workflows is a crucial issue. One way to improve these workflows is to accelerate the image acquisition procedures by fully automating the patient setup. This paper describes a system that performs this task without the use of markers or other prior assumptions. It returns metric coordinates of the 3-D body shape in real-time for inverse positioning. This is achieved by the application of an emerging technology, called Time-of-Flight (ToF) sensor. A ToF sensor is a cost-efficient, off-the-shelf camera which provides more than 40,000 3-D points in real-time. The first contribution of this paper is the incorporation of this novel imaging technology (ToF) in interventional imaging. The second contribution is the ability of a C-arm system to position itself with respect to the patient prior to the acquisition. We are using the 3-D surface information of the patient to partition the body into anatomical sections. This is achieved by a fast two-stage classification process. The system computes the ISO-center for each detected region. To verify our system we performed several tests on the ISO-center of the head. Firstly, the reproducibility of the head ISO-center computation was evaluated. We achieved an accuracy of (x: 1.73±1.11 mm/y: 1.87±1.31 mm/z: 2.91±2.62 mm). Secondly, a C-arm head scan of a body phantom was setup. Our system automatically aligned the ISO-center of the head with the C-arm ISO-center. Here we achieved an accuracy of ± 1 cm, which is within the accuracy of the patient table control.

1 Introduction and Related Work

Workflow optimization is an important task in clinical procedures. The combination of increased life expectancy and advancements in the field of medicine

* The authors gratefully acknowledge funding of the International Max-Planck-Research-School for Optics and Imaging (IMPRS-OI) and the Erlangen Graduate School in Advanced Optical Technologies (SAOT) by the German National Science Foundation (DFG) in the framework of the excellence initiative. We also thank the Siemens AG, Health Care Sector, Forchheim, Germany for supporting this work.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 549–556, 2009.

have resulted in a constantly increasing number of patients. This in-turn creates a heavier workload for hospitals and healthcare systems in general. Thus, it is becoming progressively more important to develop efficient healthcare procedures without any decline on the quality of patient care. In the future, either healthcare costs will increase dramatically or smart cost-effective solutions for optimizing current workflows have to be found.

With that last goal in mind we examined the image acquisition workflows, in general, and analyzed them in terms of bottlenecks. In this paper we introduce a system to shorten the probably most important and for sure most time consuming part of this procedure: the initial patient positioning step. Image acquisition procedures have constantly been optimized within the past years. Today the scanning time for CT and C-arm systems is basically negligible. What is still very time consuming within the workflow of image acquisition procedures is the patient-dependent setup procedure. Each patient is physiologically relatively unique. There is also a diverse number of clinical procedures. As a result, the patient setup process cannot be easily generalized. Nevertheless, the correct positioning of the patient is a crucial parameter for the quality of image acquisition. Also for retrospective hybrid-image acquisitions it is important to acquire images at the correct position. Therefore, it is important that patient positioning is performed accurately, which can make it a very time consuming task. Especially in interventional procedures correct positioning of the image acquisition device is mandatory. Todays C-arm systems are already very intuitive to use. Strobel et al. [1] describe the setup procedure and the associated time consuming steps. Several fluoroscopic images have to be taken before the required body part is within the ISO-center of the C-arm. These steps are both very time consuming and applying additional radiation doses to the patient.

To automate and speed-up the setup procedure for interventional image acquisition procedures we suggest a system based on a ToF sensor. The proposed system partitions the whole body into several 3-D bounding boxes. It operates marker-lessly and does not rely on any other assumptions. The C-arm system receives exact metric coordinates from the proposed system which can then be used in automatically adjusting the C-arm's position with respect to the patient and for data acquisition. We call this *inverse positioning*, as the system is positioned with respect to the target and not the target with respect to the system. Furthermore, the C-arm knows exactly the bounds of the object to scan and an instantaneous collision detection can be provided. Last but not least, the proposed solution is very cost-effective.

The importance of a solution to these problems is also reflected by prior work. Grimson et al. [2] introduced a system wich supports optical tracking of patient and instrument locations using surface data. Navab et al. [3] proposed a system for intraoperative positioning and repositioning of mobile C-arms using a camera-augmented mobile C-arm. This system speeds up the whole procedure and also reduces the radiation of the patient as it decreases the number of images which have to be acquired for positioning the patient. Using an optical camera and X-ray/optical markers the system supports guidance for C-arm repositioning.

Fig. 1. Examples for ToF sensor models. On the left side: CamCube from PMDtec GmbH, Germany. On the right side: SR4000 from MESA Imaging, AG, Switzerland (images taken from manufactor websites)

C-arm systems also suffer from their narrow field-of-view. Wang et al. [4] suggest a stitching algorithm for scaning long bones using a C-arm system. They, too use an additional optical camera to augment different views. Ladikos et al. [5] provide a method for collision detection for a C-arm environment. An array of optical cameras is used to generate a 3-D representation of the operation room. All of these applications are either based on markers and/or use multiple cameras/systems. The proposed system is independent of markers and only needs a single sensor.

2 Materials and Methods

2.1 Time-of-Flight Sensor

ToF sensors provide a direct way to acquire 3-D surface information [6]. ToF sensors have several advantages over other 3-D surface acquisition techniques. The most promising advantage is that ToF is on its way to become a component of mass markets like consumer electronics and the automotive industry. Currently a high-end ToF sensor is available for about USD 7,000 (see Fig. 1). A target price for ToF sensors of a few hundred dollars can be expected in the near future. ToF sensors also render calibration steps, which are mandatory for stereo based systems, unnecessary. This is based on the monocular all-solid-state architecture of ToF sensors. Such an architecture also enables a high portability of the system and a variety of integration prospects in existing systems. Recent ToF sensors provide data rates up to 25 frames per second with an lateral resolution of up to 204×204 pixels. Each of these 3-D points provides precise metric information in the sensor coordinate system. Recently several systems for medical applications like respiratory motion or patient positioning using a ToF sensor were proposed [7,8]. Further information about the principle of ToF sensors can be found in Xu et al. [6]. A detailed discussion about the advantages of ToF sensors over other 3-D surface acquisition techniques can be found in Schaller et. al [7].

2.2 Body Part Detection

In this section we will give an overview of how a ToF sensor enables inverse real-time positioning. We will introduce a generalized, very fast and fairly simple

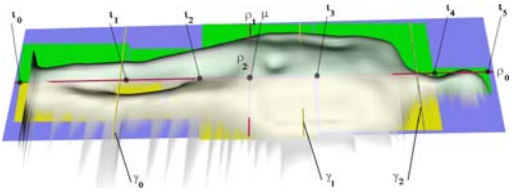


Fig. 2. Overview: The red lines indicate the three principal axes ρ_0,ρ_1,ρ_2. The three golden sections γ_0, γ_1, γ_2 are shown as yellow lines. Furthermore, all intersection points ι_k and the centroid μ are also depicted.

algorithm to solve this task. The proposed algorithm for body part detection consists of two main parts which require a preprocessing and segmentation of the ToF sensor data. The first part describes an effective and robust two-stage classification procedure. At the end of this first subtask, the 3-D body surface is subdivided into multiple anatomically meaningful regions of interest. The second part computes bounding boxes for each of these regions and their corresponding ISO-center. Furthermore, a fairly simple calibration method for aligning the computed ISO-center with the C-arm ISO-center is introduced.

The ToF sensor is rigidly mounted above the patient table and the whole patient is within the field of view of the ToF sensor. We denote $\mathbf{P}$ the $M \times N$ 3-D points of interest acquired by a ToF sensor.

$$\mathbf{P} = [\mathbf{p}_{i,j}], i \in \{0, 1, .., M-1\}, j \in \{0, 1, ..N-1\} \tag{1}$$

Typically ToF data is affected by noise. To reduce this noise we apply both, a bilateral filter [9] and a temporal averaging filter. The averaging filter returns the average 3-D point cloud using data from the last n accumulated 3-D point clouds. To reduce the amount of data and to identify 3-D points belonging to the body we detect the patient table and compute a virtual plane. For this task we use a Hough-Transform like method based on surface normals described in Schaller et al. [7]. Knowing the virtual table plane we can discard all 3-D points which do not belong to the patient. As a result for further computations only relevant 3-D points $\hat{\mathbf{P}} \subseteq \mathbf{P}$ ($\subseteq$ denotes a subset of points) belonging to the body have to be considered. To speed-up the classification process, we apply a normalization on the remaining 3-D points $\hat{\mathbf{P}}$. Therefore a Karhunen-Loeve-Transformation (PCA) is performed. As a result the origin of the coordinate system is placed in the centroid μ of the point cloud $\hat{\mathbf{P}}$ and the axes are aligned with the axial, sagittal and coronal plane of the patient. The main advantage we achieve is parallelism of the bounding boxes to the axes. This results in a more computationally efficient run-time for the algorithm. The three principal axes computed by the transformation are denoted as ρ_0, ρ_1, ρ_2 (see Fig. 2). We have prior knowledge about the shape of the human body. Therefore, we can assume that ρ_0 corresponds to the height of the patient, ρ_1 to the width and ρ_2 to the depth.

One can then compute five intersections (ι_k, where $k \in \{0, 1, .., 5\}$) along the first principle axis (see Fig. 2). The outer most intersection points ι_0 and ι_5 are

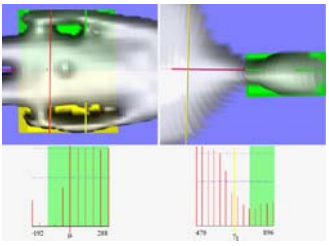


Fig. 3. Refinement of the bounding boxes. Left: Detection of the pelvic region with the corresponding histogram below. μ is the origin of the coordinate system. The search interval d_1 is 30 cm in each direction with a bining size of 4.8 cm. Right: Detection of the neck with the corresponding histogram below. γ_1 is the upper golden section intersection. The search interval d_0 is 10 cm with a bining size of 3.2 cm.

defined by the outermost 3-D points on ρ_0. Again, we utilize prior knowledge about the shape of the human body. For a coarse first stage initialization of the classification we compute three golden intersections γ_0, γ_1, γ_2 along ρ_0 (see Fig. 2). The golden section γ_2 roughly indicates the position of the neck. γ_1 is located near the diaphragm and γ_0 at the knees of the person. These points constrain the search space for the exact position of the remaining intersection points.

To refine these coarse initial values we introduce a second stage in our classification. Two histograms H_0 and H_1 along ρ_0 are computed. The vertical axis of H_0 bins the silhouette of the body along the positive p_1 axis, while the vertical axis of H_1 bins it along the positive p_2 direction. The horizontal axis of H_0 ranges from $\gamma_2 - d_0$ to $\gamma_2 + d_0$. The horizontal axis of H_1 ranges from $\gamma - d_1$ to $\gamma + d_1$. d_0 and d_1 are emirically determined. Based on these histograms, ι_4 and ι_5 are the minima of H_0 and H_1 accordingly. ι_3 is then defined as the mid-point between ι_4 and ι_5, while ι_1 is set to γ_0.

We can then use these intersections as an input for the second part of the algorithm. This part computes bounding boxes using $\iota_0, .., \iota_5$. With the exception of the head, left and right bounding boxes are computed for each body segment (e.g. left and right abdomen), where ρ_0 acts as a delimiter. In addition to the boundary values, the ISO-center and the volume is computed for each box. Figure 4 shows the full body part segmentation.

Before we are able to position a target, a fairly simple calibration step has to be performed. The following has to be done only once for the whole system. A coin is placed on a box on the patient table (see Fig. 5). The ISO-center I_{Carm} of the C-arm is manually aligned with the center of the coin. We use two fluoroscopic images, one from 0° and one from 90° to do this. After determining I_{Carm}, the corresponding 3-D coordinate (ISO-center) of the center of the coin I_{ToF} in the ToF coordinate system has to be identified. These two points, I_{Carm} and I_{ToF} can be considered as the origins of each of the corresponding coordinate systems. To position a bounding box ISO-center, the 3-D coordinate of the ISO-center is shifted to I_{ToF}.

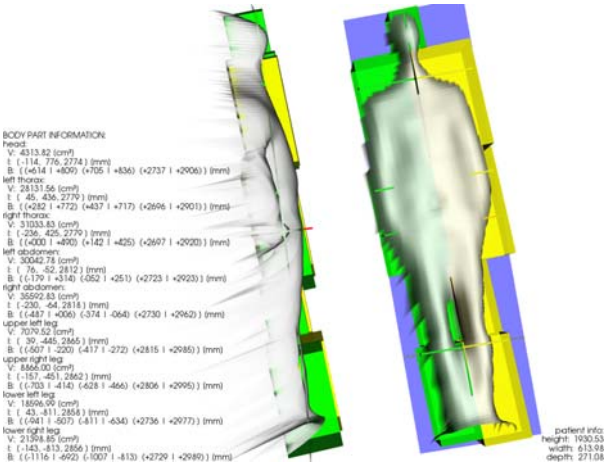


Fig. 4. Full body acquisition using a ToF sensor including bounding boxes (side view and top view). On the left side information about the different bounding boxes is shown. V shows the volume of the bounding box in cm^3, I the ISO-center of the bounding box in mm and B the bounds of the actual bounding box in mm. On the lower right side, some basic information about the dimensions of the patient is displayed.

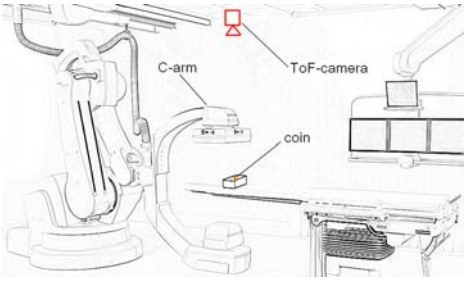


Fig. 5. Coin calibration: A coin is placed on a box on the patient table. The coin is aligned in the ISO-center of the C-arm. Afterwards, the corresponding 3-D coordinate of the position of the coin in the ToF point cloud can be determined.

3 Experiments and Results

For the evaluation we used a SR-3000 ToF sensor from MESA Imaging, AG, Switzerland and a C-arm system of the Artis zee family from Siemens AG, Healthcare Sector, Germany. We rigidly mounted the ToF sensor on the ceiling above the patient table so that the patient can be fully visible. The ToF sensor has a resolution of 176x144 pixels with a field of view of 47.5° and 39.6° respectively. The depth accuracy after preprocessing the data is below 1 mm per pixel.

We first examined the reproducibility of the ISO-center computation. Without loss of generality, this was done for the head ISO-center. For the other

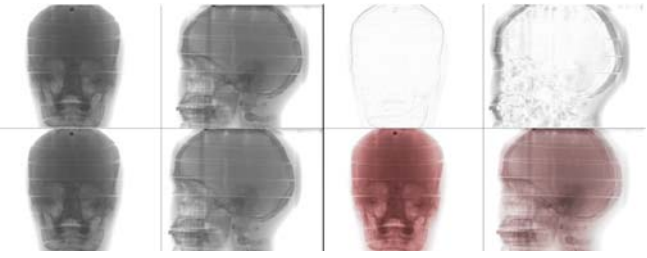


Fig. 6. Phantom evaluation: Two upper left images: 0° and 90° projections of manually aligned head (gold standard). Two lower left images: 0° and 90° projections of automatically aligned head using the proposed method. Two upper right images: 0° and 90° projection difference images of manually and automatically aligned head. Two lower right images: 0° and 90° projections of manually and automatically aligned head (automatically aligned projections are superimposed in red).

ISO-centers this could be done in a similar manner. The head ISO-center of three persons was computed 70 times per person. We took the mean x- y- z- coordinate as our reference ISO-center. We again computed 140 head ISO-centers for each of these persons while they were in the same position. The mean squared errors between those measurments and the reference in x- y- z- direction (corresponding to ρ_0,ρ_1,ρ_2) are: x: 1.73±1.11 mm, y: 1.87±1.31 mm, z: 2.91±2.62 mm. Furthermore, we computed the inter-subject standard deviation on these datasets: x: 3.16 mm y: 2.85 mm z: 4.42 mm. To evaluate our approach within a clinical setup a body phantom was placed in a C-arm environment. We were again interested in the ISO-center of the phantom head. The phantom was shifted in all three room dimensions multiple times. After computing the ISO-center I_h of the head bounding box with respect to the previously computed I_{ToF} we could directly compute the translation to align I_h with I_{Carm}. A gold standard was defined by manually aligning the ISO-center of the head using the same technique as for the previously described coin calibration (see Fig. 6). This is also the way it is done in hospitals today. We compared the table position of the gold standard alignment and the automatic alignment. In all cases we were able to automatically position the head in the x and the y direction within the accuracy of the patient table. According to the manufactor the accuracy is 1 cm. In the z-direction, we had an error of ± 1 cm (see Fig. 6) because the z-coordinate of the ISO-center heavily depends on the quality of the table segmentation. On a standard CPU (2.0 GHz dual-core) our algorithm has an execution time of 65 ms. Including data acquisition, preprocessing and segmentation, the overall execution time is about 143 ms.

4 Conclusion

Automation of time consuming steps is a key competency of future clinical procedures and workflow aware hospitals. We have shown that our system for fully

automatic patient setup is able to identify anatomical regions for image acquisition in real-time. This information can be used to align the ISO-center of anatomical targets with the ISO-center of a C-arm system fully automatically. For this task, a 3-D surface acquired by a ToF sensor is analyzed. A two-stage classification process is introduced to identify body parts and to compute the corresponding ISO-centers. The extracted information can be directly used to control any image acquisition devices. This results in an inverse positioning, where the image acquisition device can be positioned according to the patient. Furthermore, several patient dependent metric information are provided by the system. This information can be used for e.g., initial positioning, collision detection, dose estimation or respiratory motion correction [7].

References

1. Strobel, N., Meissner, O., Boese, J., Brunner, T., Heigl, B., Hoheisel, M., Lauritsch, G., Nagel, M., Pfister, M., Rúhrnschopf, E.-P., Scholz, B., Schreiber, B., Spahn, M., Zellerhoff, M., Klingenbeck-Regn, K.: 3D Imaging with Flat-Detector C-Arm Systems, 3rd edn., pp. 33–51. Springer, Heidelberg (2009)
2. Grimson, W.E.L., Leventon, M.E., Ettinger, G., Chabrerie, A., Ozlen, F., Nakajima, S., Atsumi, H., Kikinis, R., Black, P.: Clinical experience with a high precision image-guided neurosurgery system. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 63–73. Springer, Heidelberg (1998)
3. Navab, N., Wiesner, S., Benhimane, S., Euler, E., Heining, S.M.: Visual servoing for intraoperative positioning and repositioning of mobile C-arms. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 551–560. Springer, Heidelberg (2006)
4. Wang, L., Traub, J., Heining, S.M., Benhimane, S., Euler, E., Graumann, R., Navab, N.: Long bone X-ray image stitching using camera augmented mobile C-arm. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 578–586. Springer, Heidelberg (2008)
5. Ladikos, A., Benhimane, S., Navab, N.: Real-time 3D reconstruction for collision avoidance in interventional environments. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 526–534. Springer, Heidelberg (2008)
6. Xu, Z., Schwarte, R., Heinol, H., Buxbaum, B., Ringbeck, T.: Smart Pixel - Photometric Mixer Device (PMD) / New System Concept of a 3D-Imaging-on-a-Chip. In: 5th International Conference on Mechatronics and Machine Vision in Practice, pp. 259–264 (1998)
7. Schaller, C., Penne, J., Hornegger, J.: Time-of-Flight Sensor for Respiratory Motion Gating. Medical Physics 35(7), 3090–3093 (2008)
8. Schaller, C., Adelt, A., Penne, J., Hornegger, J.: Time-of-flight sensor for patient positioning. In: Proceedings of SPIE Medical Imaging 2009: Visualization, Image-guided Procedures and Modeling (2009)
9. Paris, S., Durand, F.: A fast approximation of the bilateral filter using a signal processing approach. Int. J. Comput. Vision 81(1), 24–52 (2009)

A Method to Correct for Brain Shift When Building Electrophysiological Atlases for Deep Brain Stimulation (DBS) Surgery

Srivatsan Pallavaram[1], Benoit M. Dawant[1], Rui Li[1], Joseph S. Neimat[2], Michael S. Remple[2], Chris Kao[2], Peter E. Konrad[2], and Pierre-François D'Haese[1]

[1] Dept. of Electrical Engineering and Computer Science
[2] Dept. of Neurosurgery,
Vanderbilt University, Nashville, TN, USA

Abstract. To help surgeons to pre-operatively select the target location for DBS electrodes, functional atlases based on intra-operatively acquired data have been created in the past. Recently, many groups have reported on the occurrence of brain shift in stereotactic surgery and its impact on the procedure but not on the creation of such atlases. Due to brain shift, the pre- and intra-operative coordinates of anatomic structures are different. When building large population atlases, which rely on pre-operative images for normalization purposes, it is thus necessary to correct for this difference. In this paper, we propose a method to achieve this. We show evidence that electrophysiological maps built using corrected and uncorrected data are different and that the maps created using shift-corrected data correlate better than those created using uncorrected data with the final position of the implant. These findings suggest that brain-shift correction of intra-operatively recorded data is feasible for the construction of accurate shift-corrected electrophysiological atlases.

1 Introduction

Deep brain stimulation (DBS) is a surgical procedure involving the implantation of an electrode in the deep brain to stimulate specific nuclei using a pacemaker. DBS has provided remarkable therapeutic benefits to patients suffering from movement disorders such as the Parkinson's disease. To help the surgeon pre-operatively select the target location for the electrode, functional atlases [1, 2] based on intra-operatively acquired electrophysiological data from a number of patients have been created to complement anatomical and histological atlases. Several authors [1, 3-5] have shown techniques by which such atlases can be used in the planning, placement and programming of DBS.

An underlying assumption in the creation of these atlases is that anatomical structures do not move between pre-operative imaging and intra-operative recording. A number of studies have proved that this assumption is not valid. Miyagi *et al.* [6] found the anterior and posterior commissures to be more medial, posterior and inferior on the post-operative MRI than on the pre-operative MRI. Khan *et al.* [7] reported brain shifts of up to 4 mm in deep brain structures. Using real-time intra-operative MRI Martin *et al.* [8] recently reported appreciable ipsilateral brain shift

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 557–564, 2009.

during burr hole access. Consequently, the pre- and intra-operative coordinates of anatomic structures may be different. But, to create population atlases, pre-operative image volumes are typically utilized because the stereotactic platform that is used to reference the intra-operative data is built in the pre-operative image space. It is thus critical to correct for brain shift to place data recorded intra-operatively at the correct location in the pre-operative scans. In this paper, we propose a method to do this. We also present preliminary results which indicate that shift correction has a substantial effect on statistical maps derived from these data and that shift-corrected maps are more accurate than maps computed with uncorrected data.

2 Data

With IRB approval each patient had pre-operative MRI and CT, a post-operative CT acquired on the day of the surgery, and a post-operative stable CT acquired approximately one month after surgery before programming of the electrode. The scan acquired the day of surgery will be called post-op CT, the scan acquired one month after the procedure will be called stable CT. Typical CT images were acquired at kVp = 120 V, exposure = 350 mAs and 512x512 pixels. In-plane resolution and slice thickness were respectively 0.5 mm and 0.75 mm. MRI (TR 12.2 ms, TE 2.4 ms, 256x256x170 voxels, with typical voxel resolution of 1x1x1 mm^3) were acquired using the SENSE parallel imaging technique (T1W/3D/TFE) from Philips on a 3T scanner. We used patients that underwent sub-thalamic nucleus (STN) targeting. Stimulation data included the location of each stimulation point, the efficacy (therapeutic response) observed and the associated stimulation current, the adverse effect (if any) and the associated current. Efficacy was recorded as percentage reduction in symptoms from baseline as assessed by a neurologist. Only those points with at least 70% efficacy were used. The dataset comprised of 36 efficacious points and 35 eye deviation points (an adverse effect) from 13 patients for the left side, 20 efficacious points from 8 patients and 34 eye deviation points from 12 patients for the right side. Some other adverse effects include muscular contraction, dysarthria, or parasthesia. We have chosen to use eye deviation data in this work because it is the most populous adverse effect in our dataset.

3 Method

To create atlases, the 3D T1-weighted images were registered to each other using a combination of intensity-based rigid and non-rigid registration algorithms. The non-rigid registration we proposed earlier [9] is called the Adaptive Bases Algorithm (ABA). Briefly, it computes a deformation field that is modeled as a linear combination of radial basis functions with finite support. This results in a transformation with several thousands of degrees of freedom. Two transformations (one from the atlas to the subject and the other from the subject to the atlas) that are constrained to be inverses of each other are computed simultaneously. ABA reduces the computational complexity and improves the convergence properties of related B-splines-based approaches by identifying regions of mis-registration and adapting the

compliance of the transformation locally. The algorithm arrives at the final deformation iteratively across scales and resolutions and uses mutual information as the similarity measure.

One solution to build shift-free functional atlases is to use only those patients that have minimum or no intra-operative shift (based on pneumocephalus in immediate CT) as shown by us in [10]. Patients were placed in the low, medium and large shift category based on the average air pocket width (AAPW) at the cortical surface seen on the post-op CT. To determine the region in the frontal cortex where cortical surface shift was most likely to cause shift at the target, a reference line parallel to the direction of gravity and passing through the implant was drawn and average of 4 air width measurements made in its vicinity was computed. Using data from patients in the low category (AAPW $\leq$ 3 mm) we populated shift-free atlases. These shift-free atlases are built without any correction applied to the intra-op coordinates. The limitation of this approach is that a large number of patients' data (medium: 3 mm < AAPW $\leq$ 7 mm and large group: AAPW > 7 mm) cannot be used.

The shift correction method we propose permits using all patients to build statistical maps. Fig 1(a) shows a detailed model of various shift components.

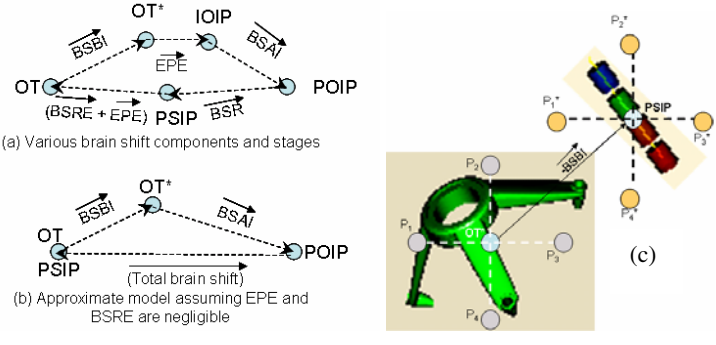


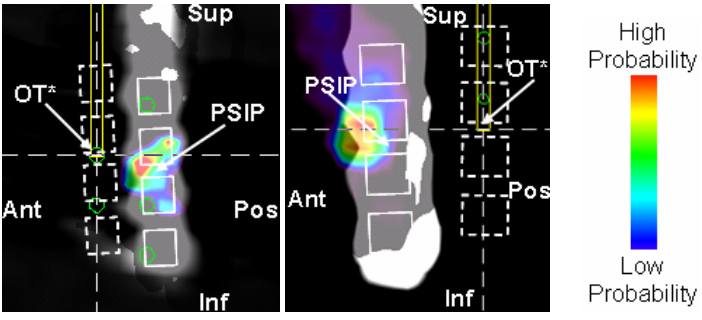
Fig. 1. (a) Detailed model of various brain shift components, (b) approximate model, (c) demonstrating how data points P_1, P_2, P_3 and P_4 are corrected for brain shift to arrive at P_1^*, P_2^*, P_3^*, P_4^* using the model in (b). OT: optimal target, $(\overrightarrow{BSBI})$: Brain Shift Before Implantation, OT*: shifted position of OT due to $(\overrightarrow{BSBI})$ and identified as the location for implantation by electrophysiological mapping, $(\overrightarrow{EPE})$: Electrode Placement Error, IOIP: Intra-Operative Implant Position, $(\overrightarrow{BSAI})$: Brain Shift After Implantation, POIP: Post-Operative Implant Position as seen on the CT acquired immediately after surgery, $(\overrightarrow{BSR})$: Brain Shift Recovery, PSIP: Post-operative Stable Implant Position as seen on the stable CT acquired about a month after surgery, $(\overrightarrow{BSRE})$: Brain Shift Recovery Error.

During the procedure, the optimal target (OT) where the implant should ideally be placed gets displaced due to Brain Shift Before Implantation $(\overrightarrow{BSBI})$ and moves to OT*. By electrophysiological mapping OT* is identified as the location for implantation. Electrode Placement Error $(\overrightarrow{EPE})$ due to the finite accuracy of the stereotactic

system and/or manual errors causes the electrode to be implanted at a slightly different location. This is the Intra-Operative Implant Position (IOIP). Brain Shift After Implantation ($\overrightarrow{BSAI}$) causes the implant to move further to the Post-Operative Implant Position (POIP) as seen on the post-op CT. Finally, due to Brain Shift Recovery ($\overrightarrow{BSR}$) the lead stabilizes at the Post-operative Stable Implant Position (PSIP) by the time the stable CT is acquired. The difference between PSIP and OT is because of ($\overrightarrow{EPE}$) and/or Brain Shift Recovery Error ($\overrightarrow{BSRE}$) due to the brain not recovering exactly to its pre-operative state. If we assume that ($\overrightarrow{BSRE}$) and ($\overrightarrow{EPE}$) are negligible, the model reduces to the simpler formulation shown in fig 1(b) where PSIP returns to OT. Using this model we can account for brain shift when populating atlases. The intra-operative coordinates of all the data points (P_1, P_2, P_3 and P_4 in fig 1(c)) and OT^* are known in the platform coordinates. Probabilistic maps built using these coordinates are referred to as uncorrected maps. Using the approximate model, the transformation between OT^* and PSIP is $-\overrightarrow{BSBI}$. By applying this transformation to all intra-operative points, their corrected coordinates (P_1^*, P_2^*, P_3^*, P_4^*) can be computed as illustrated in fig 1(c). By populating electrophysiological atlases using these coordinates we can build shift-corrected atlases and maps referred to as corrected maps. Probabilistic maps were created from atlases of intra-operative measurements using the method proposed by us in [3].

4 Results

Our method is based on the assumption that, in the stable CT, brain has recovered to its pre-operative state. This assumption is difficult to prove. However, if it is correct, the lead position in the stable CT should correspond to a region of high efficacy predicted by our shift-free map. To verify this hypothesis we have used two patients in which we observed substantial shift (3.14 mm and 2.90 mm at the target position). Fig 2 illustrates our results. In both patients, we show the location identified intra-operative as the optimal target for implantation (OT^*) shown by 4 dashed contacts and the stable lead position (solid contacts) overlaid on shift-free efficacy maps. Since shift-free maps are the closest to ground truth i.e. maps unaffected by shift, and because as per the approximate mode the stable lead position (PSIP) is the shift-recovered position of the implant (OT^*), these results show that our assumption of negligible brain shift recovery is reasonable because PSIP correlates better than OT^* with shift-free maps. To further test our assumption, we would ideally compare shift-free with corrected and uncorrected efficacy maps. Due to the relatively small number of shift-free cases for which we have stable CTs a quantitative comparison is not possible at the current time. To address this issue we validate our approach in two steps: (1) we show that corrected and uncorrected maps are different, and (2) we show that the position of the electrode in the stable CT correlates better with zones of high efficacy predicted by the shift-corrected maps than with those predicted by uncorrected maps.



(a) (b) (c)

Fig. 2. Sagittal view showing relative positions of the intra-operative optimal target (OT*) as dotted contacts, actual lead from the stable CT (PSIP) as solid contacts overlaid on shift-free maps built for two patients (a) Patient1, (b) Patient2. (c) Color scale. The gray region extending out of the core of the map and masking the stable lead is the low probability region of the map.

To show that the maps are different, we first compute the Dice similarity coefficient (S) [11] between two sets defined in (1), (2), and (3) where $p(X, Y, Z)$ is the probability value at point (X, Y, Z) in a map.

$$\Omega_{uncorrected} = \{(X,Y,Z),\ni\ p_{uncorrected}(X,Y,Z) \geq threshold\} \quad (1)$$

$$\Omega_{corrected} = \{(X,Y,Z),\ni\ p_{corrected}(X,Y,Z) \geq threshold\} \quad (2)$$

$$S = \frac{2 * N(\{\Omega_{uncorrected} \cap \Omega_{corrected}\})}{N(\Omega_{uncorrected}) + N(\Omega_{corrected})} \quad (3)$$

$$N(\Omega) \text{ is the number of elements in any set } \Omega$$

We then compare connectivity and spread of the maps. Because physiological regions should concentrate into well connected tight clusters, we compare the maps by computing the number of clusters (NOC) with high probability. We do this using 3-dimensional 26-connectivity where voxels sharing any of their 6 faces, 12 edges or 8 corners are grouped into the same cluster. A larger number of such clusters indicates greater number of isolated high probability zones and poor overall clustering. To measure the spread of the maps, we computed the mean distance from cluster centroid (MDCC) for points of high probability. Table 1(a) gives the values of dice coefficient for the uncorrected and corrected maps of efficacy (EFF) and eye deviation (EYE). Table 1(b) shows the number of 26-connected clusters in the two maps for high probability points and the mean distance from cluster centroid for those points. This table shows that the uncorrected and corrected maps are very different. From table 1(a), the uncorrected and corrected efficacy and eye deviation maps for the the left and right sides had respectively 57%, 44%, 57% and 46% of non-zero probability voxels overlapping. Considering only voxels with high probability (p >= 0.7), the percentage of overlapping voxels decreased to 0%, 12%, 0% and 8% respectively indicating that the high probability regions in the two maps were almost completely different. Table 1(b) shows that the number of distinct high probability clusters is generally higher in the uncorrected maps than in the corrected maps; 4 to 1 for left efficacy, 8 to 1 for right

eye deviation and 2 to 1 for left eye deviation. This indicates that clustering tends to be tighter in the corrected maps than in the uncorrected maps. In turn, this suggests that corrected maps represent electrophysiology better than uncorrected maps. This is further supported by the larger mean distances from cluster centroid for uncorrected maps than for corrected maps. Fig 3 and fig 4 show coronal, axial and sagittal slices of (a) corrected and (b) uncorrected maps overlaid over the stable CT for the two large-shift patients used earlier. Since we showed in fig 2 that shift recovery error can be neglected, it means that the stable lead position or PSIP returns to the optimal target location in a patient. Therefore, if shift-corrected maps built for a patient were correct then they would correlate better than uncorrected maps with PSIP. This can be seen in fig 3 and fig 4.

Table 1. (a) Dice coefficient (S) for uncorrected (uncorr) and corrected (corrr) maps, EFF: efficacy, EYE: eye deviation, (b) Number of clusters (NOC) in the two maps for p >= 0.7 and 26-connectivity, and, mean distance from cluster centroid (MDCC) for p >= 0.7

(a)	S	
	p > 0	**p >= 0.7**
EFF LEFT	0.57	0
EFF RIGHT	0.44	0.12
EYE LEFT	0.57	0
EYE RIGHT	0.46	0.08

(b)	**Map**	**NOC**	**MDCC**
EFF LEFT	**Corr**	1	1.29
	Uncorr	4	3.25
EFF RIGHT	**Corr**	2	2.5
	Uncorr	2	2.8
EYE LEFT	**Corr**	1	1.27
	Uncorr	2	1.67
EYE RIGHT	**Corr**	1	1.9
	Uncorr	8	3.64

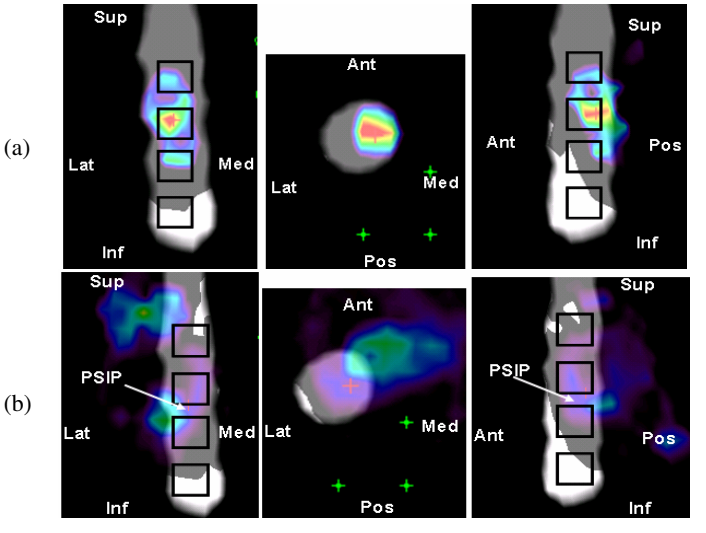


Fig. 3. Coronal, axial and sagittal slices of (a) corrected and (b) *uncorrected* maps overlaid over the stable CT for patient1. The color scale is shown in fig 2(c).

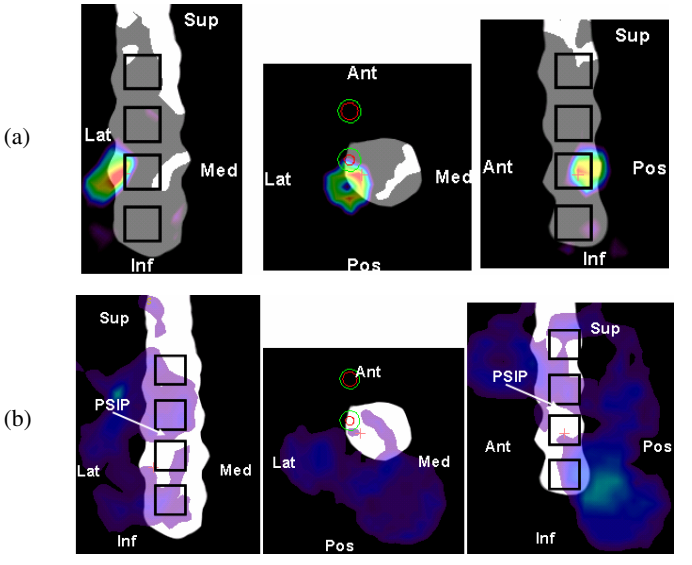


Fig. 4. Coronal, axial and sagittal slices of (a) *corrected* and (b) *uncorrected* maps overlaid over the stable CT for patient2. Color scale is shown in fig 2(c)

5 Discussion

Creation of accurate efficacy and side effect maps is important for the pre-operative planning and intra-operative guidance of DBS procedures. Stimulation data acquired during the procedures is a rich source of information to build these maps, but substantial brain shift (up to 4 mm) in areas surrounding the regions of implantation has been reported in the literature. Because of this shift, the intra-operative coordinates of a structure may be different from its pre-operative coordinates. In turn, this affects the accuracy of maps derived from these data. It has been reported that shift is related to the amount of air entering the cranial cavity during the procedure. A straightforward approach to building the maps would thus be to screen patients and keep only those for whom there is minimum air invasion. While feasible, this approach also severely limits the number of data sets which can be used to create statistical maps. In this paper, we present an approach that permits the correction of this shift and thus the use of data that would otherwise need to be discarded. Albeit preliminary, the results we present strongly suggest that our approach produces efficacy maps that correlate better with the anatomical location selected as optimal during the procedure than the uncorrected maps. This optimal anatomical target location is referred to as OT, which according to our approximate model is the stable lead position (PSIP). Work is ongoing to validate these results in a large number of patients and to correlate high efficacy zones predicted using shift-corrected maps of intra-operative data with post-operative programming stimulation response observations.

Acknowledgment. This research has been supported, in parts, by NIH R01 EB006136.

References

1. D'Haese, P.-F., Cetinkaya, E., Konrad, P.E., Kao, C., Dawant, B.M.: Computer-Aided Placement of Deep Brain Stimulators: From Planning to Intraoperative Guidance. IEEE Trans. on Medical Imaging. 24(11), 1469–1478 (2005)
2. Finnis, K.W., Starreveld, Y.P., Parrent, A.G., Sadikot, A.F., Peters, T.M.: Three dimensional database of subcortical electrophysiology for mage-guided stereotactic functional neurosurgery. IEEE Trans. on Medical Imaging 22(11), 93–104 (2003)
3. Pallavaram, S., D'Haese, P.-F., Kao, C., Yu, H., Remple, M., Neimat, J.S., Konrad, P.E., Dawant, B.M.: A new method for creating electrophysiological maps for DBS surgery and their application to surgical guidance. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 670–677. Springer, Heidelberg (2008)
4. D'Haese, P.-F., Pallavaram, S., Yu, H., Spooner, J., Konrad, P.E., Dawant, B.M.: Deformable Physiological Atlas-Based Programming of Deep Brain Stimulators: A Feasibility Study. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 144–150. Springer, Heidelberg (2006)
5. Guo, T., Finnis, K.W., Parrent, A.G., Peters, T.M.: Development and application of functional databases for planning deep-brain neurosurgical procedures. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. Part I. LNCS, vol. 3749, pp. 835–842. Springer, Heidelberg (2005)
6. Miyagi, Y., Shima, F., Sasaki, T.: Brain shift: an error factor during implantation of deep brain stimulation electrodes. Neurosurgery 107, 989–997 (2007)
7. Khan, M.F., Mewes, K., Gross, R.E., Škrinjar, O.: Assessment of Brain Shift Related to Deep Brain Stimulation Surgery. Stereotact Funct Neurosurgery 86, 44–53 (2008)
8. Martin, A.J., Larson, P.S., Ostrem, J.L., Sootsman, W., Talke, P., Weber, O.M., Levesque, N., Myers, J., Starr, P.A.: Placement of deep brain stimulator electrodes using real-time highfield interventional magnetic resonance imaging. Magn. Reson. Med. 54, 1107–1114 (2005)
9. Rhode, G.K., Aldroubi, A., Dawant, B.M.: The adaptive bases algorithm for intensity-based nonrigid image registration. IEEE Trans. On Medical Imaging 22(11), 1470–1479 (2003)
10. Pallavaram, S., Remple, M., Neimat, J.S., Kao, C., Konrad, P.E., Dawant, B.M., D'Haese, P.-F.: Effect of brain shift on the creation of functional atlases for deep brain stimulation surgery. In: Computer Assisted Radiology and Surgery, Germany (2009)
11. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology 26, 297–302 (1945)

Asymmetric Image-Template Registration

Mert R. Sabuncu[1], B.T. Thomas Yeo[1], Koen Van Leemput[1,2,3], Tom Vercauteren[4], and Polina Golland[1]

[1] Computer Science and Artificial Intelligence Lab, MIT
[2] Department of Radiology, Harvard Medical School
[3] Dep. of Information and Computer Science, Helsinki University of Technology
[4] Mauna Kea Technologies, Paris, France

Abstract. A natural requirement in pairwise image registration is that the resulting deformation is independent of the order of the images. This constraint is typically achieved via a symmetric cost function and has been shown to reduce the effects of local optima. Consequently, symmetric registration has been successfully applied to pairwise image registration as well as the spatial alignment of individual images with a template. However, recent work has shown that the relationship between an image and a template is fundamentally asymmetric. In this paper, we develop a method that reconciles the practical advantages of symmetric registration with the asymmetric nature of image-template registration by adding a simple correction factor to the symmetric cost function. We instantiate our model within a log-domain diffeomorphic registration framework. Our experiments show exploiting the asymmetry in image-template registration improves alignment in the image coordinates.

1 Introduction

Dense spatial correspondences among multiple images can be established by either directly registering pairs of images [1,2], or through a common coordinate system, where each image is aligned with a template [3,4,5]. Standard off-the-shelf pairwise registration algorithms are commonly used by both approaches. Yet, as we discuss in this paper, the two formulations are inherently different. In particular, we show that employing an algorithm designed for pairwise image registration can be sub-optimal if one of the images is a template.

In pairwise image registration, we seek a bi-directional mapping between the spatial coordinates of two images. It is therefore natural to assume that both coordinate systems should play equivalent roles in the algorithm. This symmetry of the pairwise image registration problem is exploited by inverse consistent algorithms [1,6,7,8,9,10,11]. These methods use the gradients of both images in the optimization while constraining the warps between the images to be inverses of each other, either explicitly or by construction. Empirically this strategy yields more robust and accurate alignment results [7,11].

In contrast, the problem of image alignment to a template is fundamentally asymmetric: registration must be unidirectional for the template to represent

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 565–573, 2009.

a valid probabilistic model [12,13]. In this paper, we further argue that this asymmetry is important in the context of atlas-based segmentation because the alignment quality should be measured in the image coordinate frame, rather than that of the atlas. To the best of our knowledge, this paper includes the first practical demonstration of this argument.

We present an energy-based formulation of the asymmetric image-template registration. Many of the derivations we include have previously appeared, cf. [8,10], but only in the context of *symmetric* pairwise image registration. The main point of the present paper is to demonstrate that accounting for the *asymmetry* in image-template registration will lead to improved segmentation. We develop an algorithm that employs a stationary velocity field parametrization of diffeomorphisms [11] to solve the proposed optimization problem. We include two sets of experiments on images from the OASIS [14] and Brainweb [15] datasets.

2 Problem Formulation

Given two images $I, J : \mathbb{R}^3 \mapsto \mathbb{R}$, the transformation $\Phi : \mathbb{R}^3 \mapsto \mathbb{R}^3$ that aligns the images is commonly determined via a regularized optimization problem, such as:

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmin}} \, \|I - J \circ \Phi\|_2^2 + \operatorname{Reg}(\Phi), \tag{1}$$

where $[J \circ \Phi](x) = J(\Phi(x))$ and $\|f\|_2^2 = \int_{\mathbb{R}^3} (f(x))^2 dx$ for any $f : \mathbb{R}^3 \mapsto \mathbb{R}$.

While we only consider the L_2 norm in this paper, the following discussion extends to other image dissimilarity measures. It can be shown that the L_2 norm is not symmetric with respect to spatial transformation, and the optimization problem of Eq. (1) depends on which image is warped [8,1,10]. In the absence of a preference, it makes sense to symmetrize the dissimilarity measure either by using a symmetric differential form as in [10] or by seeking

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmin}} \, \|I - J \circ \Phi\|_2^2 + \|I \circ \Phi^{-1} - J\|_2^2 + \operatorname{Reg}(\Phi) + \operatorname{Reg}(\Phi^{-1}). \tag{2}$$

Eq. (2) is the basis of many inverse consistent image registration approaches [6,7,8,9,11]. The use of symmetry results in robust and accurate optimizers that employ gradients of both images.

As discussed in [12], a true probabilistic formulation requires the model (template) to be warped and thus registration of an image should be formulated uni-directionally. This argument can be extended to energy-based approaches. In segmentation, for example, one estimates an anatomical label for each voxel in image I. A typical approach involves registering the image I with template T, which might represent the average of training images. The goal is to bring the underlying hidden labels of the image into spatial agreement with the labels of the training data. Segmentation quality depends heavily on this accordance.

Since segmentation accuracy is measured as the agreement between estimated and ground truth labels *in the spatial coordinates of the image* I, using a dissimilarity term measured in the image coordinates should improve label alignment.

This points to the "forward" formulation, where the template T is warped:

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmin}} \int_{\mathbb{R}^3} [I(x) - T(\Phi(x))]^2 \, dx + \operatorname{Reg}(\Phi). \tag{3}$$

By introducing $y = \Phi(x)$, we arrive at an equivalent formulation:

$$\hat{\Phi} = \underset{\Phi}{\operatorname{argmin}} \int_{\mathbb{R}^3} \left[I(\Phi^{-1}(y)) - T(y)\right]^2 \det(\nabla \Phi^{-1}(y)) dy + \operatorname{Reg}(\Phi), \tag{4}$$

where $\nabla \Phi^{-1}$ denotes the Jacobian matrix of transformation Φ^{-1} with respect to the spatial coordinates. Eq. (4) involves warping the image I, and hence can be viewed as a "backward" formulation. Combining Eq. (3) and Eq. (4), we achieve a "bi-directional" formulation for image-template registration:

$$\begin{aligned} \hat{\Phi} = \underset{\Phi}{\operatorname{argmin}} \frac{1}{2} \int_{\mathbb{R}^3} & [I(x) - T(\Phi(x))]^2 \\ & + \left[I(\Phi^{-1}(x)) - T(x)\right]^2 \det(\nabla \Phi^{-1}(x)) dx + \operatorname{Reg}(\Phi). \end{aligned} \tag{5}$$

Note that Eqs. (3), (4) and (5) are *theoretically identical* in the continuous domain, but may yield different numerical solutions after discretization. In particular, solving the bi-directional optimization problem of Eq. (5) will involve the gradients of both the image and template. Thus, we expect that similar to inverse consistent approaches, this formulation will enjoy robust and accurate performance. If we use an inverse consistent warp regularization, i.e., $\operatorname{Reg}(\Phi) = \operatorname{Reg}(\Phi^{-1})$, the only difference between Eq. (5) and the symmetric formulation of Eq. (2) is the Jacobian weighting term that encodes local areal distortion due to the warp. Specifically, $\det(\nabla \Phi^{-1}(x))$ is larger than 1 in regions where an area in image I is mapped to a smaller area in template T and less than 1 elsewhere. This weighting reflects the asymmetry in the objective function that evaluates alignment *in image coordinates*.

3 Log-Domain Diffeomorphic Registration

We demonstrate and compare the algorithms that solve Eq. (3), Eq. (4) and Eq. (5) in the log-domain diffeomorphic registration framework [11], which employs a Demons-style approach [16] to decouple the optimization of the similarity measure and of the regularization term, by introducing an auxiliary transformation Γ. In each iteration, the algorithm minimizes the following energy:

$$E(I, T, \Phi, \Gamma) = \operatorname{Dissim}(I, T, \Phi) + \lambda \operatorname{Dist}(\Phi, \Gamma)^2 + \operatorname{Reg}(\Gamma), \tag{6}$$

where $\lambda > 0$ and $\operatorname{Dist}(\Phi, \Gamma)$ encodes the distance between the two transformations. $\operatorname{Reg}(\Gamma) + \lambda \operatorname{Dist}(\Phi, \Gamma)^2$ replaces $\operatorname{Reg}(\Phi)$ from the previous section. In each iteration, the Demons algorithm optimizes Eq. (6) in two steps. First, we fix Γ and estimate Φ that minimizes the first two terms. Then, we fix Φ and minimize the last two terms with respect to Γ. The second step searches for a smooth approximation to the current warp Φ and typically reduces to smoothing Φ.

In the log-domain diffeomorphic Demons framework [11], a warp Γ is parameterized with a smooth, stationary velocity field $v : \mathbb{R}^3 \mapsto \mathbb{R}^3$ via an ODE: $\frac{\partial \Gamma(x,t)}{\partial t} = v(\Gamma(x,t))$ and an initial condition $\Gamma(x,0) = x$. The solution of the ODE, $\Gamma(x) = \exp(v)(x)$, can be computed efficiently using scaling and squaring [17] and inverted by using the negative of the velocity field: $\Gamma^{-1} = \exp(-v)$. An update velocity field u computed in the first step of the current iteration, can be then combined with the previous warp estimate $\Gamma = \exp(v)$ to compute the current warp $\Phi = \exp(v') \approx \exp(v + u)$ [11]. The distance between the two warps $\mathrm{Dist}(\Phi, \Gamma)$ is defined as $\|u\|_2^2$.

We now derive the update formula for the asymmetric bi-directional dissimilarity cost of Eq. (5). For a fixed deformation $\Gamma = \exp(v)$, the first step of the Demons algorithm seeks u that minimizes the following objective function:

$$\sum_{x \in \Omega} \Bigg[\Big[I(x) - T(\exp(v+u)(x)) \Big]^2 + \Big[I(\exp(-v-u)(x)) - T(x) \Big]^2 \det(\nabla \exp(-v)(x)) \Bigg] + 2\lambda \|u(x)\|^2,$$

where, for simplicity, we approximate $\det(\nabla \exp(-v-u)(x)) \approx \det(\nabla \exp(-v)(x))$ for a small u, removing dependencies between the voxels at each update and replace the integral with a discrete sum. This non-linear least squares problem can be efficiently solved using the Gauss-Newton method, which requires finding the gradient of the terms with respect to the stationary velocity field u at $u = 0$ and solving a linearized least-squares problem. We define: $\boldsymbol{g}_T(x) = \nabla T(\exp(v)(x))$, $\boldsymbol{g}_I(x) = \nabla I(\exp(-v)(x))$, $c_v(x) = \det(\nabla \exp(-v)(x))$, and

$$H(x) = \Big[\boldsymbol{g}_T(x) \boldsymbol{g}_T(x)^T - c_v(x) \boldsymbol{g}_I(x) \boldsymbol{g}_I(x)^T + 4\lambda \mathrm{Id}_{3\times 3} \Big].$$

The image gradients are column vectors, $c_v(x)$ is a scalar, $H(x)$ is a 3×3 matrix, $(\cdot)^T$ denotes the transpose and $\mathrm{Id}_{3\times 3}$ is the 3×3 identity matrix. Solving the linearized least-squares problem yields the update for each voxel x

$$u(x) = H(x)^{-1} \Big[\Big(I(x) - T(\Gamma(x)) \Big) \boldsymbol{g}_T(x) - \Big(I(\Gamma^{-1}(x)) - T(x) \Big) c_v(x) \boldsymbol{g}_I(x) \Big]. \quad (7)$$

Since the cost function in Eq. (5) is the average of the ones in Eq. (3) and Eq. (4), corresponding updates can be similarly derived. Furthermore, the symmetric formulation used by inverse-consistent registration algorithms for pairwise image alignment can be obtained by setting the Jacobian determinant term $c_v(x) = 1$ in our framework. The following is a summary of the algorithm. $\star\, K_{\sigma_K}$ denotes spatial convolution with an isotropic Gaussian with standard deviation σ_K.

Asymmetric bi-directional log-domain image-template registration:

- Initialize $\Gamma = \exp(0)$
- Iterate until convergence:
  - Given current $\Gamma = \exp(v)$, compute update $u(x)$ based on Eq. (7).

- Update: $v \leftarrow v + u$
- Regularize: $v \leftarrow v \star K_{\sigma_K}$

The algorithm has two input parameters: λ that controls the step size taken in the update (via $H(x)$) and σ_K that controls the warp smoothness.

4 Experimental Results

Using two experiments, we compare five algorithms:

1. **Template Warp**: Optimize Eq. (3).
2. **Image Warp**: Optimize Eq. (4).
3. **Asymmetric Bi-directional**: Optimize Eq. (5).
4. **Image Warp without Jacobian**: Ignore the Jacobian term in the Image Warp Algorithm and treat the template as a second image.
5. **Symmetric Bi-directional**: Ignore the Jacobian term in the Asymmetric Bi-directional Algorithm.

For each ordered pair of images, we ran all five algorithms, treating the first image as the image I and the second image as the template T. The only algorithm invariant to the ordering of the two images is the symmetric Algorithm 5.

We quantify the quality of alignment in the image coordinates using: (1) image mean square error (MSE), obtained by averaging the squared differences between the image intensities and the interpolated template intensities over the image voxels, and (2) the Dice score [18] to measure overlap of regions of interest. Dice scores were computed in the image coordinates by measuring the overlap between the image labels and the transferred template labels obtained through nearest neighbor interpolation. Dice scores vary from 0 to 1 with higher values indicating better alignment. Both MSE and Dice have been extensively used in the literature to evaluate registration results, cf. [19]. Yet, it is important to note that different applications might require different evaluation metrics.

Due to the arbitrary tradeoff between the image-based term and regularization in the objective function, comparisons between registration algorithms are fair only when considered over a range of harmonic energy of the warps [20]. Harmonic energy measures the non-linearity of the warp, defined to be the average Frobenius norm of the Jacobian of the displacement field associated with the warp. A wider kernel used to estimate Φ (larger σ_K) results in smoother warps, yielding a lower harmonic energy.

We ran each algorithm on each ordered pair of images ten times with σ_K values at equally spaced intervals from one to ten voxels and $\lambda = 10^{-3}$. For each ordered pair of subjects and each algorithm, we thus obtained 10 alignment scores with corresponding harmonic energies. We then computed the mean and standard error across trials for a particular harmonic energy.

In the two experiments, we provide a pairwise comparison of the Image Warp and Aysmmetric Bi-directional Algorithms (Algs. 2, 3) to their counterparts that ignore the Jacobian term (Algs. 4, 5). This pairwise comparison aims to investigate the effect of ignoring the Jacobian term in the cost function.

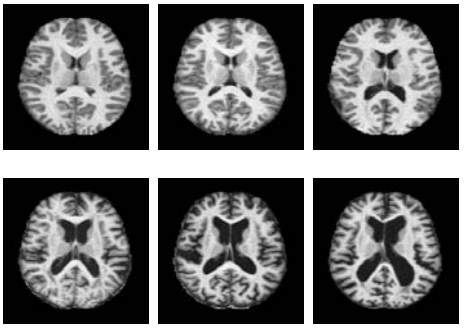


Fig. 1. OASIS: Brain MRIs of six subjects

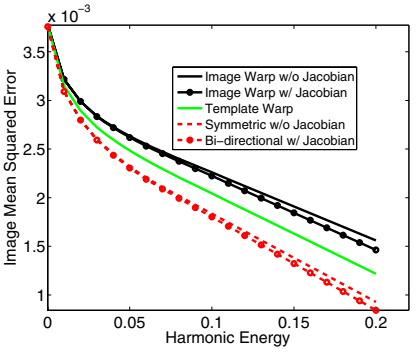


Fig. 2. OASIS: Average Image Mean Squared Error (MSE). The dynamic range of image intensities is $[0, 1]$.

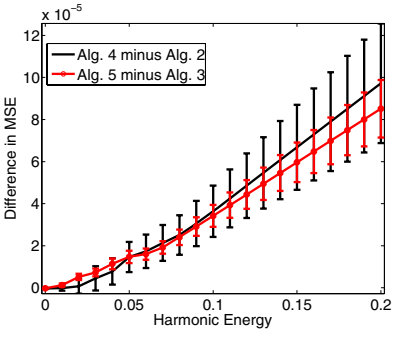


Fig. 3. OASIS: Average Image Mean Square Error (MSE) Difference. Positive values indicate that ignoring the Jacobian term increases the average error in registration.

4.1 Oasis Experiment

In the first experiment, we employed pre-processed (skull stripped and gain-field corrected) MRI volumes ($176 \times 208 \times 176$, 1 *mm* isotropic) of 6 subjects from the OASIS database [14]. The subjects represent a wide spectrum of the population, both in terms of age and dementia score. Axial slices are shown in Fig. 1, demonstrating large anatomical variability across the six subjects. We applied the five algorithms to all thirty ordered pairings of these subjects.

Fig. 2 shows the image mean squared errors (MSE) over a range of harmonic energy values averaged across all pairings. Error bars are not included to avoid clutter. It is clear that the bi-directional methods (Algs. 3 and 5) outperform the unidirectional methods (Algs. 1, 2, and 4).

Furthermore, taking into account the asymmetry of image-template registration outperforms the respective algorithms that treat both images the same way (Algs. 2 vs 4 and Algs. 3 vs 5). Fig. 3 highlights this comparison. For each ordered pair of images and each harmonic energy value, we subtracted the MSE value of

Alg. 2 (3) from Alg. 4 (5) and averaged across all pairings. Positive values indicate that the former algorithm outperforms the latter. It is clear that the algorithms incorporating the Jacobian term are statistically significantly better than the corresponding algorithms that do not. Moreover, the difference between the algorithms grows with increasing harmonic energy because under more nonlinear warps, certain regions shrink or grow more, resulting in an increased asymmetry due to substantial variations in the Jacobian map over space.

4.2 Brainweb Experiment

We used ten synthetic scans obtained from Brainweb [15]. Each scan contained a simulated T1-weighted MRI volume ($256 \times 256 \times 181$, 1 mm isotropic) with corresponding tissue label maps for CSF, gray and white matter. We ran the five algorithms on 20 random ordered pairings of the subjects. A comparison of the image mean squared error (MSE) values over a range of harmonic energy values (not shown) revealed a ranking of the algorithms consistent with the OASIS dataset: bi-directional methods outperform uni-directional methods, and taking into account the asymmetry of image-template registration improves alignment.

To further evaluate the algorithms, we used the tissue labels to quantify the overlap between the underlying tissue map of the image and interpolated tissue map of the warped template. Fig. 4a compares the Dice scores for the three tissue types of the Image Warp (Alg. 2) with its counterpart that ignores the Jacobian term (Alg. 4). Similar to the previous experiment, positive values indicate that taking the Jacobian term into account yields significantly improved alignment of the underlying tissue labels, suggesting better segmentation quality in a template-based segmentation application. Moreover, the difference between the two methods grows with increasing harmonic energy. Fig. 4b shows that this conclusion is also true for the bi-directional Algorithms (Algs. 3 and 5). Note

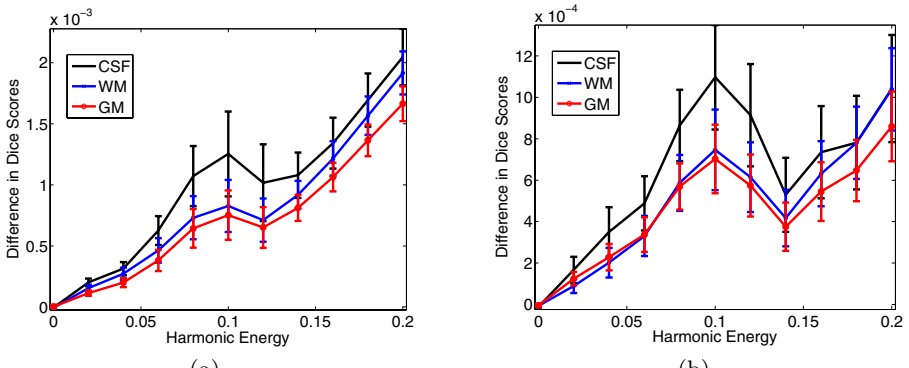


Fig. 4. Brainweb. Dice Score Differences between: (a) Image Warp with Jacobian (Alg. 2) minus Image Warp without Jacobian (Alg. 4). (b) Asymmetric (Alg. 3) minus Symmetric Bi-directional (Alg. 5).

that a Dice score difference of 10^{-3} roughly corresponds to 100 voxels in white matter, 600 voxels in grey matter and 1000 voxels in CSF.

5 Conclusions

This paper investigates the asymmetry in image-template registration. Due to this asymmetry, our analysis and experiments suggest that a straightforward application of off-the-shelf pairwise registration algorithms may be sub-optimal in the context of image-template registration. We present a bi-directional objective function that takes the asymmetry into account by introducing a correction factor: the Jacobian that quantifies the deformation of the spatial grid. This correction factor can be used to modify virtually *any* symmetric registration algorithm into an asymmetric bidirectional one.

Our experiments confirm that bi-directional methods yield improved results compared to unidirectional algorithms since the use of both image gradients in the optimization improves robustness against local minima. Furthermore, incorporating the Jacobian weighting term to account for the asymmetry produces significantly improved alignment scores in the image coordinates. This improvement becomes more pronounced under more nonlinear warps.

Acknowledgments. Support for this research is provided in part by: NAMIC (NIH NIBIB NAMIC U54-EB005149), the NAC (NIH NCRR NAC P41-RR13218), the mBIRN (NIH NCRR mBIRN U24-RR021382), the NIH NINDS R01-NS051826 grant, the NSF CAREER 0642971 grant, NCRR (P41-RR14075, R01 RR16594-01A1), the NIBIB (R01 EB001550, R01EB006758), and the NINDS (R01 NS052585-01). B.T. Thomas Yeo is funded by the A*STAR, Singapore.

References

1. Christensen, G., Johnson, H.: Consistent image registration. IEEE Transactions on Medical Imaging 20(7), 568–582 (2001)
2. Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. Neuroimage 33(1), 115–126 (2006)
3. Collins, D., Neelin, P., Peters, T., Evans, A.: Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18(2), 192–205 (1994)
4. Guimond, A., Meunier, F., Thirion, J.: Average brain models: A convergence study. Computer Vision and Image Understanding 77(2), 192–210 (2000)
5. Sabuncu, M., Balci, S., Shenton, M., Golland, P.: Image-driven population analysis through mixture-modeling. IEEE Transactions on Medical Imaging (in press, 2009)
6. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage 38(1), 95–113 (2007)

7. Avants, B., Grossman, M., Gee, J.: Symmetric diffeomorphic image registration: Evaluating labeling of elderly and neurodegenerative cortex and frontal lobe. In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 50–57. Springer, Heidelberg (2006)
8. Cachier, P., Rey, D.: Symmetrization of the non-rigid registration problem using inversion-invariant energies: Application to multiple sclerosis. In: Delp, S.L., DiGoia, A.M., Jaramaz, B. (eds.) MICCAI 2000. LNCS, vol. 1935, pp. 697–708. Springer, Heidelberg (2000)
9. Leow, A., Huang, S.C., Geng, A., Becker, J., Davis, S., Toga, A., Thompson, P.: Inverse consistent mapping in 3D deformable image registration: Its construction and statistical properties. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 493–503. Springer, Heidelberg (2005)
10. Tagare, H., Groisser, D., Skrinjar, O.: A geometric theory of symmetric registration. In: Proc. of CVPRW 2006, p. 73 (2006)
11. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Symmetric log-domain diffeomorphic registration: A demons-based approach. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 754–761. Springer, Heidelberg (2008)
12. Allassonniere, S., Amit, Y., Trouve, A.: Towards a coherent statistical framework for dense deformable template estimation. J.R. Statist. Soc.B 69, 3–29 (2007)
13. Van Leemput, K.: Probabilistic brain atlas encoding using Bayesian inference. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 704–711. Springer, Heidelberg (2006)
14. Marcus, D., Wang, T., Parker, J., Csernansky, J., Morris, J., Buckner, R.: Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of Cognitive Neuroscience 19, 1498–1507 (2007)
15. Aubert-Broche, B., Evans, A., Collins, D.: A new improved version of the realistic digital brain phantom. Neuroimage 32(1), 138–145 (2006)
16. Thirion, J.P.: Image matching as a diffusion process: An analogy with Maxwell's demons. Medical Image Analysis 2(3), 243–260 (1998)
17. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-Euclidean framework for statistics on diffeomorphisms. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 924–931. Springer, Heidelberg (2006)
18. Dice, L.: Measures of the amount of ecologic association between species. Ecology 26(3), 297–302 (1945)
19. Klein, A., Andersson, J., Ardekani, B., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G., Collins, D., Gee, J., Hellier, P., Song, J., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R., Mann, J., Parsey, R.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46(3) (2009)
20. Yeo, B., Sabuncu, M., Desikan, R., Fischl, B., Golland, P.: Effects of registration regularization and atlas sharpness on segmentation accuracy. Medical Image Analysis 12(5), 603–615 (2008)

A Demons Algorithm for Image Registration with Locally Adaptive Regularization

Nathan D. Cahill[1,2], J. Alison Noble[1], and David J. Hawkes[3]

[1] Institute of Biomedical Engineering, University of Oxford, Oxford, UK
[2] Research and Innovation, Carestream Health, Inc., Rochester, NY, USA
[3] Centre for Medical Image Computing, University College London, London, UK

Abstract. Thirion's Demons [1] is a popular algorithm for nonrigid image registration because of its linear computational complexity and ease of implementation. It approximately solves the diffusion registration problem [2] by successively estimating force vectors that drive the deformation toward alignment and smoothing the force vectors by Gaussian convolution. In this article, we show how the Demons algorithm can be generalized to allow image-driven locally adaptive regularization [3,4] in a manner that preserves both the linear complexity and ease of implementation of the original Demons algorithm. We show that the proposed algorithm exhibits lower target registration error and requires less computational effort than the original Demons algorithm on the registration of serial chest CT scans of patients with lung nodules.

1 Introduction

At a high level, image registration algorithms can be described by three components: a cost function that describes the dissimilarity between two images, a space of geometric transformations under which one or both images are allowed to deform, and, a strategy for minimizing the cost function over the space of allowable transformations. In nonparametric registration [2], the set of vector-valued functions $\{\mathbf{\Phi} : \mathbb{R}^n \mapsto \mathbb{R}^n\}$ forms the allowable space of transformations. In order to guarantee that the nonparametric registration problem is well-posed and has a smooth solution, a regularization term must be added to the dissimilarity measure to form the cost function.

In medical image registration, various regularizers, such as elastic, fluid, diffusion, and curvature have been proposed [2] for inclusion into the cost function, each defining a notion of smoothness in a slightly different way. A variety of different techniques [1,2,5,6,7] have been proposed for numerically solving the Euler-Lagrange equations arising from the variational minimization of these cost functions. In this article, we focus on Thirion's Demons algorithm [1], which is a popular choice because of its linear complexity and simple implementation.

Thirion's Demons algorithm approximates the solution to nonparametric registration with a homogeneous diffusion regularizer by iteratively performing two steps: 1) computing a *force vector* that corresponds to the variational derivative of the dissimilarity measure, and 2) convolving the force vector with a Gaussian

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 574–581, 2009.

kernel. The Demons algorithm can be generalized for use with curvature and fluid regularizers as described in [8], by reformulating the solution of the Euler-Lagrange equation as the stationary solution of a coupled system of diffusion equations.

Each of the above-mentioned regularizers are homogeneous and isotropic, meaning they ensure smoothness independently of location or direction. In the optic flow community, much research has been done into generalizing diffusion regularization to be nonhomogeneous and anisotropic [4]. Even though adaptive diffusion regularizers allow for greater flexibility, the resulting Euler-Lagrange equations are not of a form that fits nicely into the framework of an efficient, easy to implement Demons-style algorithm.

In this paper, we show that a Demons-style algorithm that uses successive Gaussian convolution *can* be constructed to handle image-driven locally adaptive regularization. This is possible if the locally adaptive regularizer is based on a generalization of the *curvature* regularizer instead of the diffusion regularizer. When the proposed locally adaptive curvature regularizer is incorporated in a variational registration problem, the resulting Euler-Lagrange equation can be linked to a coupled system of diffusion equations whose stationary solution can be approached by Demons-style successive Gaussian convolution. We illustrate the behavior of the proposed algorithm on a pedagogical example and on serial CT chest exams.

2 Background

Consider two images, a reference image R and a floating image F, both as functions in $\mathbb{R}^n$. Define a *deformation* $\mathbf{\Phi} : \Omega \subset \mathbb{R}^n \mapsto \mathbb{R}^n$ by $\mathbf{\Phi}(\mathbf{x}) = \mathbf{x} - \mathbf{u}(\mathbf{x})$, and call $\mathbf{u}$ the *displacement*. The general form of the registration problem is given by:

$$\min_{\mathbf{u}} \qquad \mathcal{E}(R, F, \mathbf{u}) := \mathcal{S}(\mathbf{u}) + \alpha \mathcal{J}(R, F^{\mathbf{u}}), \tag{1}$$

where $\mathcal{J}$ is a dissimilarity measure that quantifies the dissimilarity between the reference image R and the deformed floating image $F^{\mathbf{u}} := F(\mathbf{\Phi})$, $\mathcal{S}$ is a regularizer that ensures that the minimization problem is well-posed and that the solution is smooth in some sense, and α is a weighting parameter.

Necessary conditions for a minimum of $\mathcal{E}(R, F, \mathbf{u})$ are given by the Euler-Lagrange equations:

$$\mathcal{A}(\mathbf{u}(\mathbf{x})) = \alpha \mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x})), \tag{2}$$

with suitable boundary conditions. The partial differential operator $\mathcal{A}$ and force vector $\mathbf{f}$ arise from the Gâteaux derivatives of the regularizer and dissimilarity measure, respectively. In this article, we use the negative of the squared correlation coefficient (CC) dissimilarity measure, which yields the force vector:

$$\mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x})) = -\frac{2\rho^2_{R,F^{\mathbf{u}}}}{|\Omega|} \left[(R(\mathbf{x}) - \mu_R) - \frac{(F^{\mathbf{u}}(\mathbf{x}) - \mu_{F^{\mathbf{u}}})}{\sigma^2_{F^{\mathbf{u}}}} \right] \nabla F^{\mathbf{u}}(\mathbf{x}), \tag{3}$$

where μ_A, σ_A^2, and $\rho_{A,B}$ refer to the mean, variance, and correlation coefficient, respectively. We note that our subsequent analysis is not dependent on this choice of $\mathcal{J}$.

The diffusion and curvature regularizers [2] and their corresponding partial differential operators given by:

$$\mathcal{S}_{diff}(\mathbf{u}) = \frac{1}{2}\int_\Omega \sum_{j=1}^{n} \|\nabla u_j(\mathbf{x})\|^2 d\mathbf{x} \ , \qquad \mathcal{A}_{diff}(\mathbf{u}(\mathbf{x})) = -\Delta\mathbf{u}(\mathbf{x})\,, \tag{4}$$

$$\mathcal{S}_{curv}(\mathbf{u}) = \frac{1}{2}\int_\Omega \sum_{j=1}^{n} (\Delta u_j(\mathbf{x}))^2 d\mathbf{x}, \qquad \mathcal{A}_{curv}(\mathbf{u}(\mathbf{x})) = \Delta^2\mathbf{u}(\mathbf{x})\,. \tag{5}$$

For large deformation problems, these partial differential operators can be applied to the velocity field $\mathbf{v}(\mathbf{x},t)$ of the deformation, which is related to the displacement field by the material derivative:

$$\mathbf{v}(\mathbf{x},t) = \partial_t\mathbf{u}(\mathbf{x},t) + (\nabla\mathbf{u}(\mathbf{x},t))^{\mathrm{T}}\,\mathbf{v}(\mathbf{x},t)\,. \tag{6}$$

2.1 Locally Adaptive Diffusion Registration

Image-driven adaptive diffusion regularization can be enabled by applying a scalar weighting function to the diffusion regularizer; i.e.

$$\mathcal{S}_{diff}^{id\text{-}adapt}(\mathbf{u}) = \frac{1}{2}\int_\Omega \beta^{\mathbf{u}}(\mathbf{x}) \left(\sum_{j=1}^{n} \|\nabla u_j\|^2 \right) d\mathbf{x}. \tag{7}$$

The weighting function $\beta(\mathbf{x})$ can be chosen based on the type of behavior desired. Alvarez [3] proposes using a weighting function that is inversely proportional to the gradient magnitude of the underlying image. Charbonnier [9] and Bruhn [10] show that a function fitting this description is $\beta(\mathbf{x}) = \Psi_\epsilon\left(\|\nabla F(\mathbf{x})\|^2\right)$, where

$$\Psi_\epsilon\left(s^2\right) = \frac{\epsilon}{\sqrt{s^2+\epsilon^2}}. \tag{8}$$

Kabus [11] proposes a different type of weighting function that is based on a segmentation of the image into foreground/background regions. He also allows the weighting function to deform over the course of registration according to $\mathbf{u}(\mathbf{x})$. Here, we use the weighting function (8) but allow it to deform according to Kabus' convention.

2.2 Thirion's Demons

The first variant of Thirion's Demons algorithm [1] is essentially an approximation to the stationary solution of the following diffusion equation:

$$\begin{aligned} \partial_t\mathbf{v}(\mathbf{x},t) - \Delta\mathbf{v}(\mathbf{x},t) &= \alpha\mathbf{f}(\mathbf{x},R,F,\mathbf{u}(\mathbf{x},t))\,, \\ \mathbf{v}(\mathbf{x},0) &= 0, \end{aligned} \tag{9}$$

Algorithm 1. Demons Algorithm for Diffusion Registration

Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{v}(\mathbf{x}, t_0) = \mathbf{u}(\mathbf{x}, t_0) = 0$, $m = 0$.
repeat
  $\mathbf{v}(\mathbf{x}, t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big) * \big[\alpha \mathbf{f}(\mathbf{x}, R, F, \mathbf{u}(\mathbf{x}, t_m)) + \mathbf{v}(\mathbf{x}, t_m)\big]$.
  $\mathbf{u}(\mathbf{x}, t_{m+1}) \leftarrow \mathbf{u}(\mathbf{x}, t_m) + \tau \left[\mathbf{I} - \nabla \mathbf{u}(\mathbf{x}, t_m)\right]^{\mathrm{T}} \mathbf{v}(\mathbf{x}, t_{m+1})$.
  $m \leftarrow m + 1$.
until convergence.

where $\mathbf{x} \in \mathbb{R}^n$ and $t \geq 0$. Furthermore, the stationary solution of (9) is equivalent to nonparametric registration with a diffusion regularizer applied to the velocity field. The Demons algorithm exploits the fact that the Gaussian kernel is the Green's function of the diffusion equation, and therefore approximates the stationary solution of (9) by successive Gaussian convolution. Algorithm 1 describes this process, using the notation $\mathcal{K}\big(\mathbf{x}, \sqrt{2\tau}\big)$ to indicate a Gaussian kernel at position $\mathbf{x}$ with standard deviation $\sqrt{2\tau}$, and $*$ to indicate convolution.

Note that in Thirion's description [1] of the Demons algorithm, a normalized version of the sum of squared distances similarity measure is used. Here, we generalize to allow for any similarity measure.

3 Designing a Locally Adaptive Demons Algorithm

In order to incorporate locally adaptive regularization into a Demons-style registration framework, we must somehow relate the Euler-Lagrange equations (2) to one or more diffusion equations of the form (9). If this is possible, we can exploit the nature of the Green's function of (9) to construct an approximate solution by successive Gaussian convolution. Unfortunately, due to the nonhomogeneous structure of $\mathcal{A}_{diff}^{id\text{-}adapt}$, (which is omitted here for lack of space), there appears to be no direct way to relate the resulting Euler-Lagrange equations to diffusion equations of the form (9).

However, such a relationship *can* be established if locally adaptive *curvature* regularizers are constructed. In this section, we show how to construct an image-driven locally adaptive curvature regularizer, how to relate it to a coupled system of diffusion equations, and finally, how to exploit this relationship to design a locally adaptive Demons algorithm for registration.

3.1 Locally Adaptive Curvature Regularization

An image-driven locally adaptive curvature regularizer can be constructed in the same manner as its diffusion counterpart (7) by applying a weighting function to the homogeneous curvature regularizer. This yields:

$$\mathcal{S}_{curv}^{id\text{-}adapt}(\mathbf{u}) = \frac{1}{2} \int_{\Omega} \beta^{\mathbf{u}}(\mathbf{x}) \left(\sum_{j=1}^{n} (\Delta u_j)^2 \right) d\mathbf{x}. \tag{10}$$

When the corresponding partial differential operator $\mathcal{A}_{curv}^{id\text{-}adapt}$ is determined and applied to the velocity field, the corresponding Euler-Lagrange equation can be written as:

$$\Delta[\beta^{\mathbf{u}}(\mathbf{x})\,\Delta\mathbf{v}] = \frac{1}{2}\left(\sum_{j=1}^{n}(\Delta v_j)^2\right)\nabla\beta^{\mathbf{u}}(\mathbf{x}) + \alpha\mathbf{f}(\mathbf{x},R,F,\mathbf{u}(\mathbf{x}))\,. \tag{11}$$

3.2 Coupled PDE System

The structure of the left hand side of the Euler-Lagrange equation (11) allows us to relate the solution of (11) to the stationary solution of a coupled system of diffusion equations. This coupled system is given by:

$$\begin{aligned}\partial_t\mathbf{w}(\mathbf{x},t) - \Delta\mathbf{w}(\mathbf{x},t) &= \sqrt{\alpha}\,\mathbf{f}(\mathbf{x},R,F,\mathbf{u}(\mathbf{x},t)) \\ &\quad - \frac{\sqrt{\alpha}}{2}\left(\sum_{j=1}^{n}(w_j(\mathbf{x},t))^2\right)\nabla\left[1/\beta^{\mathbf{u}(\mathbf{x},t)}(\mathbf{x})\right], \qquad (12)\\ \partial_t\mathbf{v}(\mathbf{x},t) - \Delta\mathbf{v}(\mathbf{x},t) &= \sqrt{\alpha}\,\mathbf{w}(\mathbf{x},t)\,/\beta^{\mathbf{u}(\mathbf{x},t)}(\mathbf{x})\,, \qquad (13)\\ \mathbf{w}(\mathbf{x},0) &= \mathbf{v}(\mathbf{x},0) = 0.\end{aligned}$$

In order to establish the relationship between the stationary solution of (12)–(13) and the solution of (11), we first note that the stationary solution of (12)–(13) satisfies $\partial_t\mathbf{w} = \partial_t\mathbf{v} = 0$. Therefore, as $t \to \infty$, we see that solving for $\mathbf{w}(\mathbf{x},t)$ in (13) and substituting the result into (12) yields (11).

3.3 Locally Adaptive Demons

We have shown for the locally adaptive curvature regularizer that the solution to the Euler-Lagrange equation is equivalent to the stationary solution of a coupled system of diffusion equations. Therefore, we can exploit the structure of the Green's function of the diffusion equation in order to define a Demons-style algorithm for registration with this adaptive regularizer. Algorithm 2 lists the resulting Demons algorithm.

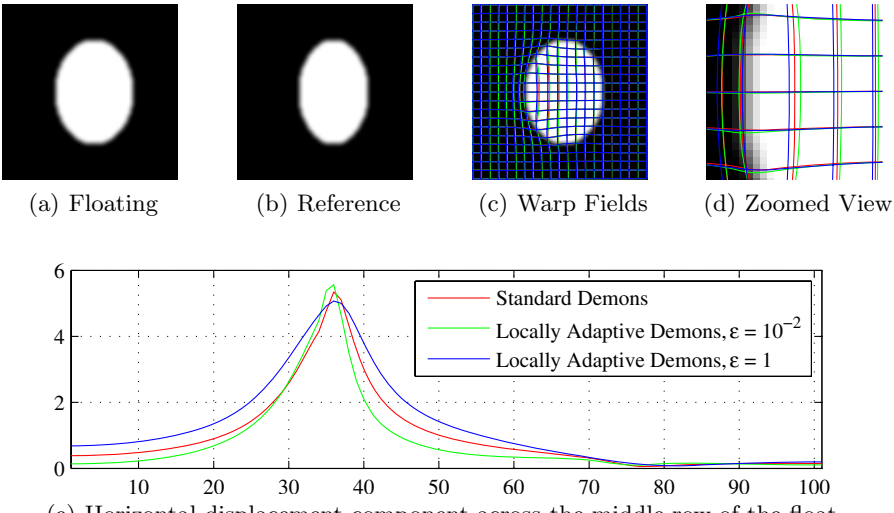
Algorithm 2. Demons Algorithm for Locally Adaptive Registration

Select time step τ; define $t_j = j\tau$, $j = 0, 1, \ldots$.
Set $\mathbf{w}(\mathbf{x},t_0) = \mathbf{v}(\mathbf{x},t_0) = \mathbf{u}(\mathbf{x},t_0) = 0$, $m = 0$.
repeat
  $\mathbf{w}(\mathbf{x},t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x},\sqrt{2\tau}\big) * \big[\,\{\text{RHS of (12) at } t = t_m\} + \mathbf{w}(\mathbf{x},t_m)\,\big]$.
  $\mathbf{v}(\mathbf{x},t_{m+1}) \leftarrow \mathcal{K}\big(\mathbf{x},\sqrt{2\tau}\big) * \big[\sqrt{\alpha}\,\mathbf{w}(\mathbf{x},t_{m+1})\,/\beta^{\mathbf{u}(\mathbf{x},t_m)}(\mathbf{x}) + \mathbf{v}(\mathbf{x},t_m)\,\big]$.
  $\mathbf{u}(\mathbf{x},t_{m+1}) \leftarrow \mathbf{u}(\mathbf{x},t_m) + \tau\,[\mathbf{I} - \nabla\mathbf{u}(\mathbf{x},t_m)]^{\mathrm{T}}\mathbf{v}(\mathbf{x},t_{m+1})$.
  $m \leftarrow m + 1$.
until convergence.

4 Pedagogical Example

To illustrate the behavior of the locally adaptive Demons algorithm, we consider a pedagogical example of registering two ellipses. Figure 1 shows two 100×100 images of an ellipse with smoothed edges; the floating ellipse has left edge that is five pixels to the left of the left edge of the reference ellipse. The right edges of both ellipses are in the same location.

In a simple experiment, we registered the ellipses using the standard and locally adaptive Demons algorithms. Figures 1(d) and 1(c) show that each algorithm warps the edge of the floating ellipse in a similar manner, but that interior of the floating ellipse is warped at different rates. Figure 1(e) affirms this point: the use of locally adaptive Demons with $\epsilon = 10^{-2}/\epsilon = 1$ causes the interior of the ellipse to be less/more affected by the ellipse edge than the standard Demons algorithm.



(a) Floating (b) Reference (c) Warp Fields (d) Zoomed View

(e) Horizontal displacement component across the middle row of the floating image. X axis of plot corresponds to pixel location.

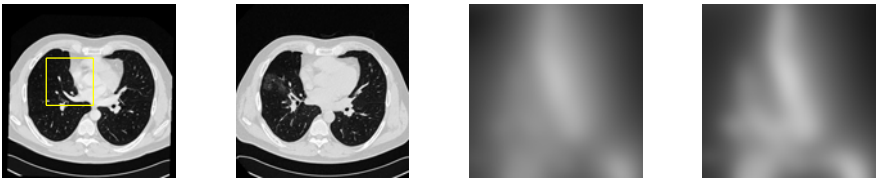
Fig. 1. Pedagogical example: warp fields correspond to standard Demons algorithm with $\alpha = 10^{-4}$(red), locally adaptive Demons algorithm with $\alpha = 10^{-4}$ and $\epsilon = 10^{-2}$ (green), and locally adaptive Demons algorithm with $\alpha = 10^{-3}$ and $\epsilon = 1$ (blue)

5 Serial Chest CT Experiment

A cursory analysis of Fig. 1(e) suggests that the locally adaptive Demons algorithm can be tuned to enable a range of behavior in how deformations are smoothed near edges in the image. To explore this behavior on a practical example, we register serial chest CT examinations of 18 patients with lung nodules.

Table 1. Statistics of TRE across all nodules, ATRE across all cases, and ECS for rigid registration and nonrigid registration via standard and locally adaptive Demons

Statistic	Rigid only	Rigid + standard Demons	Rigid + locally adaptive Demons
Mean TRE (mm)	8.5	2.6	2.5
Median TRE (mm)	7.2	2.0	1.8
Std. dev. TRE (mm)	6.5	2.2	2.1
Mean ATRE (mm)	8.4	2.3	2.1
Median ATRE (mm)	7.8	2.4	1.9
Std. dev. ATRE (mm)	4.5	1.6	1.5
ECS Percentiles (25%, 50%, 75%)		(53.0, 56.4, 56.5)	(17.2, 19.0, 21.8)



(a) (b) (c) (d)

Fig. 2. Chest CT registration: (a) Axial slice from prior volume; (b) Axial slice from current volume; (c) $\log(1 + \|\nabla \mathbf{u}\|)$ from yellow subregion of (a) using standard Demons; (d) $\log(1 + \|\nabla \mathbf{u}\|)$ from yellow subregion of (a) using locally adaptive Demons

To establish a set of ground truth points that can be used to measure target registration error (TRE), we manually identified the centers of lung nodules less than 6mm in diameter that are observable in both the prior and current images of a patient. This yielded from 4-20 ground truth points for each patient. We resampled the images to approximately $3 \times 3 \times 3\text{mm}^3$ isotropic resolution and performed a rigid preregistration step. We then performed nonrigid registration using the standard and locally adaptive Demons algorithms with various values of α and ϵ. Registration was performed in a multiresolution pyramid at four resolution levels.

Computational requirements of each algorithm were measured in terms of *effective convolution steps* (ECS). A unit of ECS is defined as the amount of computation required to convolve a vector field at the finest resolution level with a Gaussian kernel. Hence, 50 iterations of standard Demons at each of the two finest levels requires $50 + (1/8) * 50 = 56.25$ ECS. Adaptive Demons would require twice as many Gaussian convolution steps, yielding 112.5 ECS.

Table 1 shows statistics on the TRE values across all nodules, aggregate TRE (ATRE) values across all cases for the best performing algorithms (standard: $\alpha = 10^{-2}$; locally adaptive: $\alpha = 10^{-2}$, $\epsilon = 10^{-2}$), and ECS. ATRE for a particular patient case is defined as the median TRE value for all nodules in the case. Figure 2 illustrates an axial slice from the prior and current CT images of one

patient, along with images of $\log(1 + \|\nabla \mathbf{u}\|)$ for the deformations resulting from the standard and locally adaptive Demons algorithms.

6 Conclusion

We have shown how the Demons algorithm can be generalized to handle a locally adaptive regularizer. This is done by constructing an image-driven locally adaptive version of the curvature regularizer, and then relating the solution of the Euler-Lagrange equations for registration to the stationary solution of a pair of diffusion equations. The proposed algorithm is easy to implement and exhibits linear complexity in the number of pixels/voxels. Experiments on serial registration of chest CT images of patients with lung nodules show an improvement in TRE as well as a reduction in the amount of computational effort required.

References

1. Thirion, J.P.: Image matching as a diffusion process: an analogy with maxwell's demons. Medical Image Analysis 2(3), 243–260 (1998)
2. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, Oxford (2004)
3. Alvarez, L., Esclarín, J., Lefébure, M., Sánchez, J.: A PDE model for computing the optical flow. In: Proceedings XVI Congreso de Ecuaciones Diferenciales y Aplicaciones, September 1999, pp. 1349–1356 (1999)
4. Weickert, J., Bruhn, A., Brox, T., Papenberg, N.: A survey on variational optic flow methods for small displacements. In: Mathematical Models for Registration and Applications to Medical Imaging, pp. 103–136. Springer, Heidelberg (2006)
5. Bro-Nielsen, M.: Medical Image Registration and Surgery Simulation. PhD thesis, Technical University of Denmark, DTU (1996)
6. Crum, W.R., Tanner, C., Hawkes, D.J.: Anisotropic multi-scale fluid registration: Evaluation in magnetic resonance breast imaging. Physics in Medicine and Biology 50(21), 5153–5174 (2005)
7. Cahill, N.D., Noble, J.A., Hawkes, D.J.: Fourier methods for nonparametric image registration. In: Proc. CVPR Workshop on Image Registration and Fusion, June 2007, pp. 1–8 (2007)
8. Cahill, N.D., Noble, J.A., Hawkes, D.J.: Demons algorithms for fluid and curvature registration. In: Proc. International Symposium on Biomedical Imaging (June 2009)
9. Charbonnier, P., Blanc-Féraud, L., Aubert, G., Barlaud, M.: Two determinstic half-quadratic regularization algorithms for computed imaging. In: Proc. 1994 IEEE International Conference on Image Processing., vol. 2, pp. 168–172 (1994)
10. Bruhn, A., Weickert, J., Kohlberger, T., Schnörr, C.: A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. International Journal of Computer Vision 70(3), 257–277 (2006)
11. Kabus, S.: Multiple-Material Variational Image Registration. PhD thesis, der Universität zu Lübeck (October 2006)

A Meta Registration Framework for Lesion Matching*

Sharmishtaa Seshamani[1], Purnima Rajan[1], Rajesh Kumar[1], Hani Girgis[1], Themos Dassopoulos[3], Gerard Mullin[2], and Gregory Hager[1]

[1] Department of Computer Science, Johns Hopkins University, Baltimore, MD
sharmi@jhu.edu
[2] Johns Hopkins Hospital, Baltimore, MD
[3] Washington University, School of Medicine St. Louis, MO

Abstract. A variety of pixel and feature based methods have been proposed for registering multiple views of anatomy visible in studies obtained using diagnostic, minimally invasive imaging. A given registration method may outperform another depending on anatomical variations, imaging conditions, and imaging sensor performance, and it is often difficult a priori to determine the best registration method for a particular application. To address this problem, we propose a registration framework that pools the results of multiple registration methods using a decision function for validating registrations. We refer to this as *meta registration.* We demonstrate that our framework outperforms several individual registration methods on the task of registering multiple views of Crohn's disease lesions sampled from a Capsule Endoscopy (CE) study database. We also report on preliminary work on assessing the quality of registrations obtained, and the possibility of using such assessment in the registration framework.

1 Introduction

Minimally invasive diagnostic imaging methods such as flexible endoscopy, and wireless capsule endoscopy (CE) often present multiple views of the same anatomy. Duplication issues are particularly severe in the case of CE, where peristaltic propulsion may lead to duplicate information over several consecutives frames, and also several minutes apart. This may be difficult to detect, since each individual image captures only a small portion of anatomical surface due to limited working distance of these devices, providing relatively little spatial context. Given the relatively large anatomical surfaces (e.g. the Gastrointestinal tract (GI)) to be inspected, it is important to identify duplicate information as well as to present all available views of anatomical and disease views to the clinician for improving consistency, efficiency and accuracy of diagnosis and assessment.

* Supported in part by National Institutes of Health with Grant 5R21EB008227-02 and Johns Hopkins University internal funds.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 582–589, 2009.

In particular, we are interested in automatically assessing the severity of Crohn's disease, an inflammatory condition of the GI tract using CE images (Figure 1). The aggressiveness of the course of mucosal healing treatment for Crohn's disease depends upon the severity assessment of unique lesions in the CE images. We aim to identify and assess multiple views of a selected lesion, to reduce redundant information and to improve the consistency and accuracy of disease assessment.

Fig. 1. A mild Crohn's lesion in CE. CE color Images are typically 576x576 pixels.

In the literature, lesion finding has been commonly formulated as a detection problem [1] where a classifier is trained to learn the visual properties of the chosen object category (i.e. lesions). This process typically requires feature extraction to generate a low dimensional representation of image content, followed by classifier training to distinguish the desired object model(s) [2]. For CE, appearance modelling has been used for blood detection [3,4], topographic segmentation[5] and lesion classification [6].

Generic detection is different from *matching* an instance of an object to another. Therefore, we consider detection of repetitive lesions as a registration and matching problem. A registration method evaluates an objective function or similarity metric to determine a location in the target image (a second view) where a reference view (i.e. a lesion) occurs. Once a potential registration is computed, a decision function must be applied to determine the validity of the match. Our goal is to develop a trained statistical classifier that makes a decision based on the quality of a match between two regions of interest (ROIs) or views of the same lesion, rather than the appearance of the features representing an individual ROI.

Decision functions for image matching have traditionally been designed as thresholded classifiers based on one or more registration metrics. Although such a method ensures low false positive rates, the overall retrieval rate is bounded by the recall rate of the most sensitive metric. The work of Szeliski et al. [7] and Stewart et al. [8] are examples of such problem formulations. In many cases, a single, unique global threshold may not exist; therefore the determination of an adaptive threshold is a challenging problem. An integrated classifier that distinguishes registrations based on a feature representation populated by a wide range of metrics is likely to outperform such threshold based methods. Chen et al. [9] introduce a new feature vector that represents images using an extracted feature set. However, this approach still requires the same similarity metric across the entire feature set. By contrast, we present a framework that incorporates multiple registration algorithms, a generalizable feature representation for classification and a regression based ranking system to choose the highest quality match.

2 Meta Registration Framework

Given an ROI R_i in an image I_i, and a target image I_j, we define the registration function $T(R_i, I_j) \mapsto R_j$, which generates an ROI match pair (R_i,R_j). The objective function for this registration is based upon the invariance properties of image data to be matched. For example, histogram representations are invariant to rotation, whereas pixel based methods are generally not. Feature based methods are typically less affected by changes in illumination and scale. Due to large variation in these invariance properties within endoscopic studies, a single registration method may not be adequate for retreiving matches in endoscopic data. We use multiple independent global registration methods (each more accurate in a different subset of the data) and a decision function to estimate a valid match. Let us denote the similarity metric relating visual properties of the two ROIs R_i and R_j as $d(R_i, R_j)$. Using a set of global registration functions $\mathcal{T} = T_i(R, I) : i = 1 \dots n$ and estimated matches $R'_1 \dots R'_n$ from each of these methods, the decision function D which determines which estimates are correct matches can be written as in equation 1. In the case where there are multiple correct match estimates, we use a ranking function to determine the best result. Figure 2 shows the information flow in the proposed method.

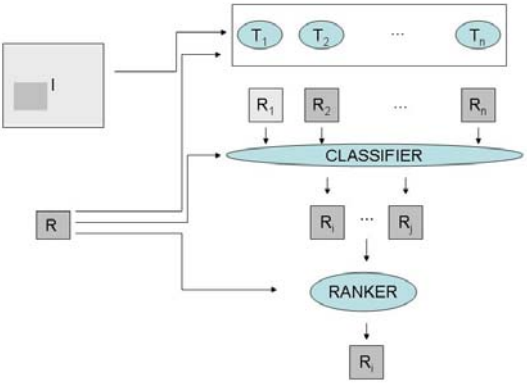


Fig. 2. Multiple independent registration methods and a robust classifier are used to determine the location of ROI R in image I

$$D(R_i, R'_j) = \begin{cases} 1, & \text{if } d(R_i, R'_j) < \gamma \\ -1 & \text{otherwise} \end{cases} \quad (1)$$

Decision Function Design. For an ROI R, we use the following notation for representing appearance features. We start with pixel based features. The intensity band of the image is denoted as R_I. The Jacobian of the image is denoted $R_J = [R_x, R_y]$ where R_x and R_y are the vectors of spatial derivatives at all image pixels. Condition numbers and the smallest Eigenvalues of the Jacobian are denoted as R_{JC} and R_{JE} respectively. The Laplacian of the image is denoted as R_{LAP}. We denote histogram based features as : R_{RGBH}, R_{WH} and R_{WCH} for RGB histograms, Gaussian weighted intensity histograms and Gaussian weighted color histograms respectively. We also consider MPEG-7 descriptors [10]: R_{EHD} (Edge Histogram), R_{Har} (Haralick Texture), R_{DCD} (Dominant Color) and R_{HTD} (Homogeneous Texture).

Given two images I_a and I_b where A is an ROI in I_A with center x and B is an ROI in I_B, we generate a feature vector for a pair of regions A and B

Table 1. Metrics Table

Metric Name	Formula
RMS (rms)	$\sqrt{(\frac{1}{n}\sum_k (A_I - B_I)^2)}$
RMS Shuffle	$\sqrt{(\frac{1}{n}\sum_k shuffle(A_I, B_I))}$
Ratio of Condition Numbers	$min(A_{JC}, B_{JC})/max(A_{JC}, B_{JC})$
Ratio of Smallest Eigenvalues	$min(A_{JE}, B_{JE})/max(A_{JE}, B_{JE})$
Laplacian Shuffle Distance	$shuffle(A_{LAP}, B_{LAP})$
Weighted Histogram Bhattacharya Distance	$sqrt(A_{WH}.B_{WH})$
RGB Histogram Bhattacharya Distance	$sqrt(A_{RGB}.B_{RGB})$
Edge Histogram Manhattan Distance	$\sum(A_{EHD} - B_{EHD})$
Haralick Descriptor Canberra Distance	$\sum \frac{\lvert A_{Har} - B_{Har}\rvert}{\lvert A_{Har}\rvert + \lvert B_{Har}\rvert}$
HTD Shuffle Distance	$shuffle(A_{HTD}, B_{HTD})$
Forward Backward check	$\lvert x - T(I_b, I_a, T(I_a, I_b, x))\rvert$

populated with the metrics shown in Table 1. The decision function can then be trained to distinguish between correct and incorrect matches using any standard classification method. We use Support Vector Machines (SVM) [11] in our experiments.

The Ranking Function: We treat the registration selection as an ordinal regression [12] problem. Given a feature set corresponding to correctly classified registrations, $F = \{f_1, ..., f_N\}$ and a set of N distances from the true registrations, we can form a set of preference relationships between the elements of F. Let us define the set of preference pairs P as, $P = \{(x, y) | f_x \prec f_y\}$. Our goal is to compute a continuous real-valued ranking function K such that, $f_x \prec f_y \in P \Longrightarrow K(f_x) < K(f_y)$. A preference pair $(x, y) \in P$ can be thought of as a pair of training examples for a standard binary classifier. We train a binary classifier C such that,

$$C(F_x, F_y) = \begin{cases} 0, & if (x, y) \in P \\ 1 & otherwise \end{cases}$$

and $C(F_y, F_x) = 1 - C(F_x, F_y)$. Given such a classifier, the rank can be computed as, $K(F) = \sum_{i=1}^{n} C(F, F_i)/n$ where K is the fraction of the training set that are less preferred to F based on the classifier. Thus K orders F relative to the training set. Let f_x represent the metrics or features of registration and $f_{i,j}$ represent the vector concatenation of f_i and f_j. The training set, $Train = \{< f_{i,j}, 0 >, < f_{j,i}, 1 > | (i, j) \in P\}$ is used to train an SVM. For classification, we pair each vector in the test set with all the vectors in the training set and use the empirical order statistics $K(F)$ described above for enumerating the rank.

3 Training Data

Given an ROI R and a set of CE images $\mathcal{I} = \{I_i : i = 1 \ldots N\}$, the task is to build a dataset of pairs of images representing correct and incorrect matches

of a global registration. We first compute the correct location of the center of the same ROI in $\mathcal{I}$ through manual selection followed by a local optimization. We denote this set of locations as $\mathcal{X} = \{X_i : i = 1 \dots N\}$. Next, we select any global registration method T and apply it between R and each image in the set $\mathcal{I}$ to generate a set of estimated ROI center locations $\mathcal{X}' = \{X'_i : i = 1 \dots N\}$ and pairs $\mathcal{R} = \{R, R_i : i = 1 \dots N\}$. The pairs are designated a class label y (correct or incorrect matches) by thresholding on the Euclidean distance (in pixels) between X_i and X'_i. We refer to this as the ground truth distance. The training set $\mathcal{T}$ contains all registered pairs and their associated class labels.

4 Experiments

Our CE study repository contains studies acquired with an approved Institutional Review Board (IRB) protocol. A CE study may contain up to 8 hours of data acquired at 2 images per second. The CE study database contains selected annotated images containing Crohn's Disease (CD) lesions manually selected by our clinical collaborators. These provide the ROIs we use for our experiments. A lesion may occur in several neighboring images, and these selected frames form a lesion set. Figure 3 shows an example of a lesion set. In these experiments, 150x150 pixel ROIs were selected. Various lesion sets contain between 2 and 25 image frames. Registration pairs were then generated for every ROI in the lesion set, totaling 266 registration pairs.

We used the folowing five standard techniques for 2D registration. These include SIFT feature matching, mutual information optimization, weighted histograms (grayscale and color) and template matching. For each of these methods, we performed a registration to estimate a registered location, resulting in a total of 1330 estimates (5 registration methods per ROI-image pair). The ground truth for these estimates was determined by using a threshold of 25 pixels on the Euclidean distance described in Section 3. The dataset contained 581 correct (positive examples) and 749 incorrect (negative examples) registrations.

For every registration estimate, we computed the registered ROI for the training pair. The feature vector respresenting this registration estimate was then computed as described in Section 2. We then trained the decision function using all registration pairs in the dataset. The performance of this integrated classifier was evaluated using a 10-fold cross-validation. Figure 5(a) shows the performance of our classifier, including comparison with the ROC curves of individual

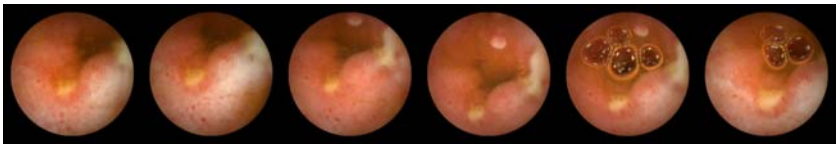


Fig. 3. A Lesion Set: a set of neighboring CE images with the same lesion

Table 2. 8 Iterations of Ranking Experiment

	Iter1	Iter2	Iter3	Iter4	Iter5	Iter6	Iter7	Iter8
No: of pairs	300	600	900	1200	1500	1800	2100	2400
Train mis-classification rate	0.001	0.014	0.016	0.015	0.018	0.017	0.017	0.017
Train pair mismatch rate	0.16	0.18	0.17	0.16	0.16	0.16	0.16	0.15
Test pair mismatch rate	0.32	0.38	0.32	0.26	0.38	0.32	0.35	0.27
Test rank mean	0.53	0.69	0.55	00.35	0.69	0.55	0.61	0.44
Test rank std dev	0.14	0.15	0.20	0.28	0.19	0.23	0.21	0.29

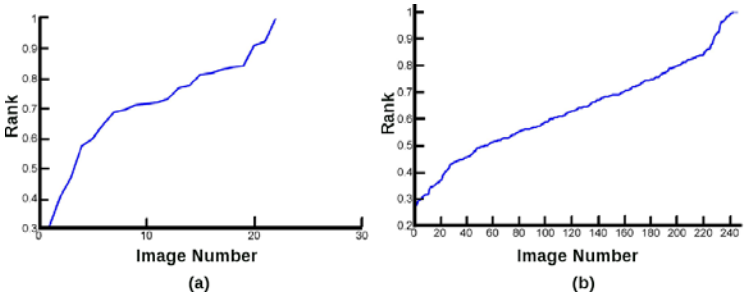


Fig. 4. Estimated Ranks for training (left) and test(right) datasets

metrics used for feature generation. The true positive rate was 96 percent and the false negative rate was 8 percent.

For n registrations, a total of n choose 2 preference pairs can be generated. We used a subset of this data as the input to the ranking model. Features used to generate a training pair include the difference between EHDs and the difference between the DCDs. Training was initiated with a random selection of $n = 300$. This estimate was then improved by iteration and addition of preference pairs at every step. At each iteration, a 10-fold cross validation was performed where the dataset was divided into appropriately sized training and test sets. Training was then conducted using an SVM model with a radial basis kernel.

Once this classifier was trained, preference relationships were predicted by classifying vectors paired with all training vectors. Relative ranks within each set were determined and pair mismatch rates were calculated. A mismatch is any pair of registrations where $K(F_x) > K(F_y)$ and $F_x \prec F_y$ or $K(F_x) < K(F_y)$ and $F_x \succ F_y$. The training mis-classification rate is the percentage of contradictions between the true and predicted preference relationships in the training set. Table 2 shows the rank metrics for each iteration. Figure 4 shows the sorted ranking function of training and test set registrations at the end of 8 iterations.

The registration framework was applied to all 266 image pairs. For each pair, all five registration methods were applied to estimate matching ROIs. The first row of Table 3 shows the number of correct registrations evaluated using the ground truth distance. Features were extracted for all registrations and the integrated classifier trained above was applied. A leave one out cross-validation

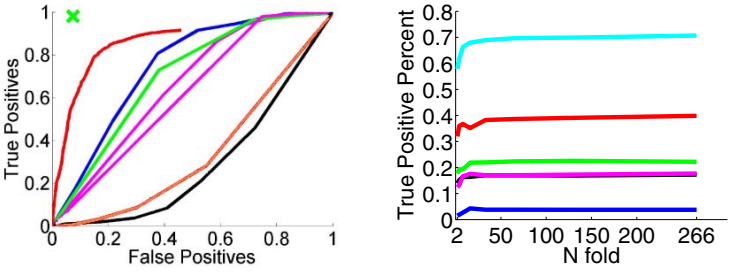


Fig. 5. Meta Registration Results. LEFT: ROC Curves of all metrics used individually overlaid with the integrated classifier (Green X), RIGHT: Percentage true positives for cross validation for template matching(red), SIFT (green), Mutual Information (blue), Intensity Weighted Histogram matching (black) Color Weighted Histogram matching (magenta) and meta registration (cyan).

Table 3. Meta Registration Results. The first row shows the number of correct matches qualified by a manual validation. The second row shows the number of correct matches qualified by the automated integrated classifier. The third row shows the number of automatically qualified matches that match the manual (ground truth) validation.

Type	Template Matching	Sift	Mutual Info	Intensity Weighted Histogram	HSV Weighted Histogram	Meta Registration
Ground Truth	165	122	54	111	129	266
Classifier	129	62	25	75	77	188
True Positives	106	59	10	46	47	188

was performed for each ROI-image pair. The second row of Table 3 shows the number of matches that the classifier validated as correct. Finally, the last row shows the number of true positives which is the number of correctly classified matches that are consistent with the ground truth classification. The last column in the table shows the performance of the meta registration. The new framework retreived a larger number of registrations than any single registration method. We performed a range of n-fold validations on the same dataset for n ranging from 2-266 (where $n = 2$ divides the set into two halves and $n = 266$ is the leave one out validation). Figure 5(b) shows the percentage of true positives retreived (which is the ratio of true positives of the meta registration to the number of correct ground truth classifications) by each individual registration method and the integrated classifier (top curve in cyan). The meta registration was able to retrieve 70 percent of matches where the best registration method only retreived 39 percent.

5 Discussion

We have proposed a meta registration framework for matching of lesions in capsule endoscopic video. This general approach incorporates multiple independent optimizers and an integrated classifier combined with a trained ranker to select the best correct match from all registration results. Our results show the integrated method outperforms the use of any single method for ROI matching and retreival in CE imagery. In future work, we plan to extend this to matching ROIs without explicit application of global optimizers.

References

1. Taylor, C.J., Cooper, D.H., Graham, J.: Training models of shape from sets of examples. In: Proc. of the British Machine Vision Conference, pp. 9–18. Springer, Heidelberg (1992)
2. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 511–518 (2001)
3. Jung, Y.S., Kim, Y.H., Lee, D.H., Kim, J.H.: Active blood detection in a high resolution capsule endoscopy using color spectrum transformation. In: Proc. of the International Conference on BioMedical Engineering and Informatics, Washington, DC, USA, pp. 859–862. IEEE Computer Society, Los Alamitos (2008)
4. Li, B., Meng, M.Q.H.: Computer-based detection of bleeding and ulcer in wireless capsule endoscopy images by chromaticity moments. Computers in Biology and Medicine 39(2), 141–147 (2009)
5. Cunha, J.P., Coimbra, M.T., Campos, P., Soares, J.M.: Automated topographic segmentation and transit time estimation in endoscopic capsule exams. IEEE Transactions on Medical Imaging 27(1), 19–27 (2008)
6. Bejakovic, S., Kumar, R., Dassopoulos, T., Mullin, G., Hager, G.: Analysis of crohns disease lesions in capsule endoscopy images. In: Proc. of the International Conference on Robotics and Automation, pp. 2793–2798 (2009)
7. Szeliski, R.: Prediction error as a quality metric for motion and stereo. In: Proc. of the International Conference on Computer Vision, Washington, DC, USA, p. 781. IEEE Computer Society, Los Alamitos (1999)
8. Yang, G., Stewart, C., Sofka, M., Tsai, C.L.: Registration of challenging image pairs: Initialization, estimation, and decision. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(11), 1973–1989 (2007)
9. Chen, X., Cham, T.J.: Learning feature distance measures for image correspondences. In: Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, pp. 560–567. IEEE Computer Society, Los Alamitos (2005)
10. Coimbra, M., Cunha, J.: Mpeg-7 visual descriptors contributions for automated feature extraction in capsule endoscopy. IEEE Transactions on Circuits and Systems for Video Technology 16(5), 628–637 (2006)
11. Vapnik, V.N.: The nature of statistical learning theory. Springer-Verlag New York, Inc., New York (1995)
12. Herbrich, R., Graepel, T., Obermayer, K.: Support vector learning for ordinal regression (1999)

Automatic Robust Medical Image Registration Using a New Democratic Vector Optimization Approach with Multiple Measures

Matthias Wacker[1,2] and Frank Deinzer[2,3]

[1] Institute of Medical Statistics, Computer Sciences and Documentation, Friedrich-Schiller-University of Jena, GER
matthias.wacker@mti.uni-jena.de
[2] Siemens AG, Healthcare Sector, Forchheim, GER
[3] University of Applied Sciences Würzburg-Schweinfurt, GER
frank.deinzer@fhws.de

Abstract. The registration of various data is a challenging task in medical image processing and a highly frequented area of research. Most of the published approaches tend to fail sporadically on different data sets. This happens due to two major problems. First, local optimization strategies induce a high risk when optimizing nonconvex functions. Second, similarity measures might fail if they are not suitable for the data. Thus, researchers began to combine multiple measures by weighted sums. In this paper, we show severe limitations of such summation approaches. We address both issues by a gradient-based vector optimization algorithm that uses multiple similarity measures. It gathers context information from the iteration process to detect and suppress failing measures. The new approach is evaluated by experiments from the field of 2D-3D registration. Besides its generic character with respect to arbitrary data, the main benefit is a highly robust iteration behavior, where even very poor initial guesses of the transform result in good solutions.

1 Introduction

Image registration can be described as the task of finding a transformation from the coordinate system of one image into another in a way that corresponding information is aligned [1]. This challenging task is a highly frequented area of research and subject of numerous publications in the last 15 years [2,3,4,5]. Dependend on the application ranging from neuro surgery to radio therapy and many more, the requirements include a high accuracy, time efficiency and robustness which are of course highly conflictive.

State of the art approaches use similarity measures that try to model the quality of the registration, given a certain transformation. Based on this objective function an optimization scheme is applied to search for the best transform. So there exist two main reasons for a failed registration. First, the optimization algorithm may get stuck in a wrong local optimum. This problem can be handled up to a certain extent by more complex search strategies at the cost

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 590–597, 2009.

of computational time [6]. Second, similarity measures can simply fail, what is often overseen. It means that the measure has its extremum at a location in the parameter space that does not correspond to meaningful registration. Although Penney et al. found already in 1998 [3] that different measures can lead to different results, this problem has hardly been further considered.

In order to avoid the problem of failing measures, researchers constructed more complex ones [7] and started to build convex combinations of two or more similarity formulations [4]. Feldmann et al. [8] used the idea of democratic integration of Triesch et al. [9], where an arbitrary number of measures are used simultaneously in a dynamically weighted sum. Since summation approaches have severe limitations as they suffer from scaling issues, it can be said that the overall approach to tackle that aspect has not been found yet.

In this paper we design a new iterative vector optimization scheme that will allow for an efficient and robust registration. It works with a set of similarity measures that are all treated equally so that the highly different scaling of the single measures does not affect their influence compared to the remaining ones. For detecting failing measures, the necessary context information can be extracted from the iteration trajectory and is used to suppress their influence.

First of all, section 2 will give a motivation of the algorithm design, followed by the formal definition. Secondly, section 3 demonstrates the capabilities of the method in two experiments. The paper closes with a discussion in section 4.

2 Method

In general, similarity measures are highly nonconvex functions with unpredictable response behavior to parameter changes. Therefore, they are hard to normalize. Although it is mostly possible to give theoretical lower and upper bounds, their actually used range is not known in advance. But we know by design that all similarity measures try to characterize the point of optimal registration what implies that they should share an extremum.

The working hypotheses of the presented approach assumes the following situation. A subset of the similarity measures works quite well, once the parameters are close enough to the solution. The rest of the measures fails to give useful information. These subsets depend on the data set and dynamically on the current guess of the parameters which may be out- or inside the region of convergence for some measures.

2.1 Motivation

For a start, we give a brief demonstration about the problem of summation approaches. Figure 1 shows three different functions on the left side and their sum on the right side. Two of the three functions share an optimal point - the sought solution. Beneath the functions, arrows denote the iteration directions $d(f_i)$ a gradient descent method would suggest. On the right side, the failing function obviously dominates in the sum $d(f)_\Sigma$. But instead of building the

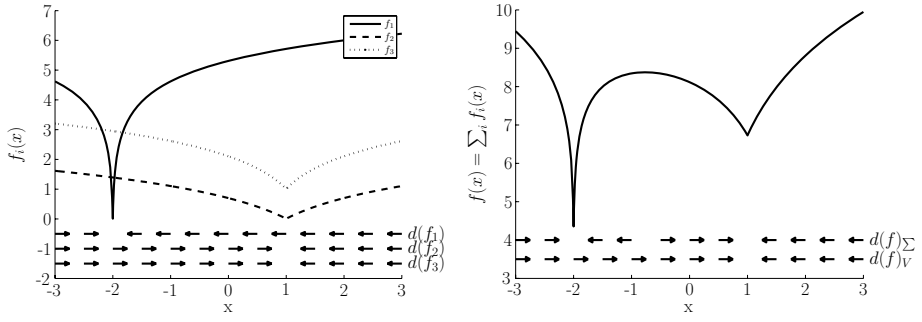


Fig. 1. Iteration directions of a gradient descent approach on single measures and a sum of measures compared to voting

gradient of the sum we could take the similarities as vector valued information. The gradient of each component of this vector votes for an iteration direction to build a majority decision. The result $d(f)_V$ is shown in the second row of arrows on the right side. Note that the algorithm would iterate right through the wrong minimum to the correct solution. Be also aware of the fact, that a multiplication of the wrong function with an arbitrary large number would still give the same voting result while the summation approach would be lost.

Considering n-dimensional problems, we recognize that there is an infinite set of directions to vote for. Of course, this is an ill posed task. Hence, we have to reinterpret the voting of a direction candidate as an averaging of the direction suggestions. The bad side effect of this averaging is that failing measures have a disturbing effect. But assuming we know which measures currently fail, we can easily suppress their influence by a small weighting factor. In terms of optimization, this failing can easily be described: let the measure be a function that has to be minimized, and let us assume we have already a previous iteration. If the last iteration was locally an ascend direction for the measure, it would have preferred the opposing direction and by that voted against the chosen one. This simple observation will be the core idea of the following method.

2.2 Registration by Vector Optimization

In the field of vector optimization, *Edgeworth-Pareto (EP)* optimality is defined with respect to an ordering cone [10]. In general, the set of EP-optimal points is of infinite cardinality. To define a single solution, we incorporate the working hypothesis by a diagonal matrix I with rank deficiency. It selects the working measures which share an optimal point $\hat{\boldsymbol{x}}$ and by that, characterize a strong EP-optimal point for this reduced sub problem. Associating with $\hat{\boldsymbol{x}}$ the maximum rank of I (namely $\operatorname{rank}(\hat{\boldsymbol{x}})$), for which $\hat{\boldsymbol{x}}$ is still a strong EP-optimal point, we write the registration problem as

$$\max_{I \in \mathbb{R}^{m \times m}} \left\{ \operatorname{rank} \left(\arg\min_{\boldsymbol{x} \in \mathbb{R}^n} f_I(\boldsymbol{x}) \right) \right\}, \text{ where } f_I(\boldsymbol{x}) := I \cdot f(\boldsymbol{x}), \;\; f(\boldsymbol{x}) \in \mathbb{R}^m. \quad (1)$$

2.3 Algorithm Design

Our iterative optimization algorithm that realizes (1) will repeat the same basic steps as classical approaches: determine an iteration direction $\boldsymbol{d}$ and a step width λ. The presented method will use a scale parameter δ that gets reduced when oscillating behavior occurs. As it is straightforward, we omitted the embedding in a multiresolution approach.

Iteration Direction. To compute the iteration direction, we approximate gradient information by central differences with evaluations of the functions in a distance of δ. To create a democratic approach where each measure f_i has an equal influence independent from its scaling, we reduce the gradient to directional information as a vector with unit length. Without loss of generality we assume that all measures have to be minimized. The continuous equivalent for the weighted discrete vote idea can be formalized as weighted averaging of the negated normalized gradients

$$\boldsymbol{d}(\boldsymbol{x}) = -\frac{1}{N}\sum_i w_i \nabla_{\|\cdot\|} f_i(\boldsymbol{x}) \quad , \quad \text{where } \nabla_{\|\cdot\|} f_i(\boldsymbol{x}) := \frac{\nabla f_i(\boldsymbol{x})}{\|\nabla f_i(\boldsymbol{x})\|}. \tag{2}$$

For the moment, the weights w_i should be considered as given, because they will be described after presenting the framework.

Step Size. In order to determine a step size, usually a one dimensional optimization is applied. A weighted sum of all measures would be a possible objective function for that. But the above example already revealed the disadvantages of scalar valued approaches. In account of that, we define a rough but robust method: the scale parameter δ of the algorithm suggests two candidates for the step size $\lambda_1 = \delta$ and $\lambda_2 = 2 \cdot \delta$. Between the two candidates, a discrete weighted vote is performed, that is summing up the weights of the measures that would prefer λ_1 and comparing it with the corresponding value of λ_2 where the larger value marks the result.

Adaption of the Weight Factors. Changing the weights in (2) during optimization results in a dynamically varying shape of the ordering cone realized by the linear mapping I in (1). Adapting the weights is the most crucial task in the algorithm. So we have to design an automatism that makes the weighting strategy meet our requirements. In this approach, the weights are generated by multiplication of two terms

$$w_i = w_{i,RE} \cdot w_{i,CO} \quad , \tag{3}$$

that are described in the following.

Reliability of a Measure ($w_{i,RE}$). To detect failing measures, we observe over time how often a measure gives a useful contribution to the iteration. Therefore, we define a history of length h to realize a time limited context extraction from the current iteration t_c to t_{c-h}. Within this context we determine to which extent the iteration direction was a descend direction for the measure. This can be formalized as having an angular of less than $\frac{\pi}{2}$ from the negated gradient

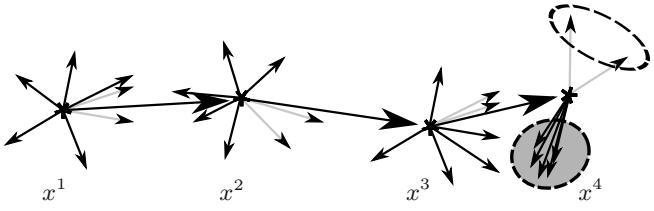


Fig. 2. Problem of inertia of reliability weights

direction to the iteration direction. A normalization is applied to keep the weights in $[0, 1]$. The angular between vectors a and b is denoted by $\sphericalangle(a, b)$

$$w^1_{i,RE} = \frac{\#\{t|\sphericalangle(-\nabla_{\|\cdot\|} f_i^t, d^t) < \frac{\pi}{2}, t \in \{t_{c-1}, t_{c-2}, \ldots, t_{c-h}\}\}}{\sum_j \#\{t|\sphericalangle(-\nabla_{\|\cdot\|} f_j^t, d^t) < \frac{\pi}{2}, t \in \{t_{c-1}, t_{c-2}, \ldots, t_{c-h}\}\}} \quad . \tag{4}$$

As a random similarity measure could get relative high weight, we penalize volatile gradient changes with a multiplication ramp. The most 'random' measure gets multiplied by one and the least random measure gets multiplied with $\gamma > 1$ to get a $[1, \gamma]$ normalization.

$$\tilde{w}^2_{i,RE} = \sum_{t_k \in \{t_c, \ldots, t_{c-h}\}} \sphericalangle(\nabla_{\|\cdot\|} f_i^{t_k}, \nabla_{\|\cdot\|} f_i^{t_{k-1}}) \tag{5}$$

$$w^2_{i,RE} = \gamma - (\gamma - 1) \cdot \frac{\tilde{w}^2_{i,RE} - \min_j \tilde{w}^2_{j,RE}}{\max_l \tilde{w}^2_{l,RE} - \min_j \tilde{w}^2_{j,RE}} \tag{6}$$

Together with $w^1_{i,RE}$ we get the total reliability in the given time context

$$w_{i,RE} = w^1_{i,RE} \cdot w^2_{i,RE} \tag{7}$$

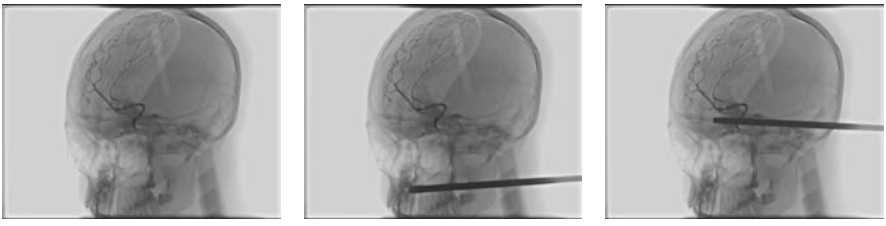
Conformity of a Measure ($w_{i,CO}$). The reliability weight can introduce a problem of inertia as the cardinal numbers in (4) can change only by one in each iteration. Figure 2 shows four following iterations $x^1, \ldots, x^4$ with the negated gradients of the measures. The preferred directions of the measures differ a lot in the first three steps. Two measures (gray) point every time in iteration direction and get therefore a high reliability weight. In the fourth iteration all except these two measures point into the same direction but the reliability could make the algorithm iterate in the opposing direction. Thus, we also consider the unweighted mean of negated gradients. Measures that point closely in that direction are conform with the majority in the current situation.

$$w_{i,CO} = \frac{\pi - \sphericalangle\left(\nabla_{\|\cdot\|} f_i^{t_c}, \sum_\nu \nabla_{\|\cdot\|} f_\nu^{t_c}\right)}{\sum_j \pi - \sphericalangle\left(\nabla_{\|\cdot\|} f_j^{t_c}, \sum_\nu \nabla_{\|\cdot\|} f_\nu^{t_c}\right)} \quad . \tag{8}$$

Subtraction from the maximum angle π and division gives again a $[0, 1]$ normalization of the weights.

3 Experiments and Results

The used data was acquired from three different phantoms (skull, pelvis and thigh) with a Siemens AXIOM Artis dFA C-arm device. To get a rigid 2D-3D



(a) (b) (c)

Fig. 3. Image data without (a) and with (b),(c) disturbing instruments present

registration problem, we reconstructed 3D volumes of $512 \times 512 \times 392$ voxels and reduced them to half size, took images of size 1024×1024 and resampled them to a resolution of 256×256. In the following, all error values represent the mean projected distance of corresponding points (equally spaced grid points in the center $(\frac{1}{3})^3$-th volume) in mm on the detector plane. As the phantoms were not moved between reconstruction and image acquisition the system protocol could be used as ground truth information. Due to the limited precision of the C-arm device, the error values have a disturbance in the range of a few mm. As setup for the presented algorithm we chose $h = 5$, $\gamma = 2$ (6) and the following set of similarity measures: *energy of differences (ED), gradient correlation (GC), gradient differences (GD), joint entropy (JE), mutual information (MI), normalized cross correlation (NCC), pattern intensity (PI), ratio of image uniformity (RIU)* and *sum of squared differences (SSD)* (see [3,5] and references).

3.1 Robustness to Heavily Disturbing Image Contents

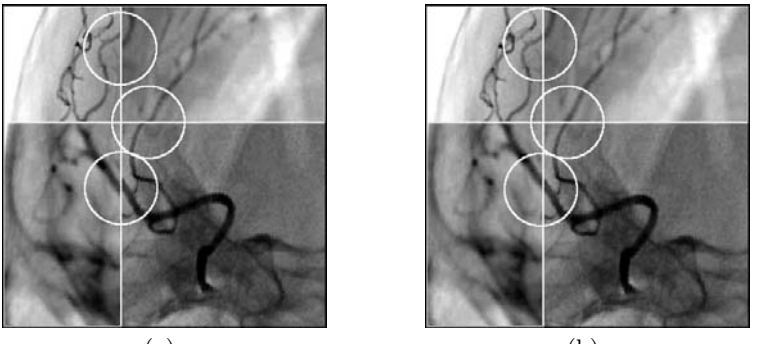
In the first experiment, we simulate surgical instruments by adding a screwdriver to the skull phantom after it has been reconstructed. Some measures are sensitive to such disturbing image content. Figure 3 shows the original phantom without and with the instrument for two different positions. We compare the presented approach with best-neighbor (BN) optimization approaches of single similarity measures. We generated 10 random initial parameter sets (3D rigid transforms) in the range of 19 mm to 42 mm distance to the optimal solution leading to the results, shown in Table 1. On this data set, MI and PI failed, gradient correlation worked quite well in the case of (a) but its error increases by almost 1 cm in the problem (b). The presented voting approach shows only a minimum increase of 0.1 mm. Although the the error values appear high, what is due to the mentioned C-arm inaccuracy, the results are very useful as it can be seen from the checker board representations of Fig. 4. Looking at the small veins in the registration result in of Fig. 4(b) we recognize an exact match.

3.2 Robustness to Bad Initial Guesses

We generated randomly 15 initial guesses and started from each of them 9 different registration problems: three different views on each of the three phantoms. The results were sorted according to their initial displacement that ranged from

Table 1. Error table of the first experiment. Error values are the medians of the error distribution of all start positions given in mm.

Problem	Voting	BN - GC	BN - MI	BN - SSD	BN - PI
3(a)	3.5	3.5	26.6	4.2	25.0
3(b)	3.6	13.3	32.5	4.6	23.1
3(c)	3.6	4.4	26.5	5.6	22.2



(a) (b)

Fig. 4. Checker board representation of the system calibration (error 0 mm) (a) and a registration result (b) which has an error value of more than 3 mm

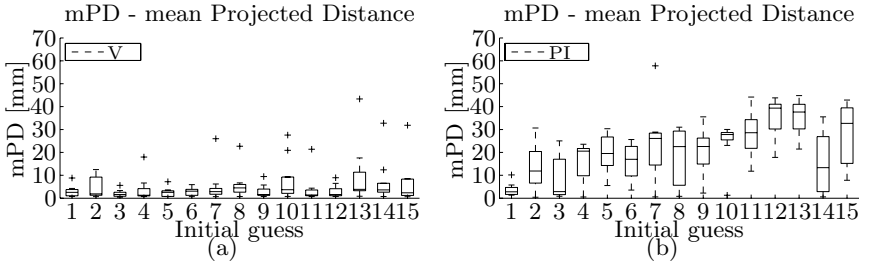


Fig. 5. Error distributions sorted by initial guesses for the vector optimization approach in (a) and a best neighbor approach with pattern intensity in (b). Boxes show 0.25-, 0.50-, 0.75-quantiles, outliers are marked as points.

10.2mm up to 43.3mm. To our knowledge, this means an unpublished wide attraction range for medical image registration. Figure 5(a) shows the resulting error distribution beginning with the closest start position '1' and ending with the farthest one '15', where each distribution represents the results of all nine registration problems. Where the voting approach shows only very few outliers without a general drift to higher values, the best neighbor approach with PI shows a high sensitivity to local optima in Fig. (b).

4 Discussion

In this paper, we formalized the problem of medical image registration by means of vector optimization (1) and introduced a new algorithm that solves the problem

accordingly. It makes use of multiple similarity measures and is therefore able to handle the problem of failing measures. Unlike other state of the art combination approaches, it avoids the summation of function values and by that the problem of single dominating similarity measures. Instead, in each iteration each similarity measure votes for an iteration direction according to its gradient information. This is a direct consequence of the working hypotheses that most of the similarity measures share a local optimum and therefore prefer the same iteration direction. The influence of each measure is controlled by dynamically adapted weights ([2]). Those use the iteration context by taking past iterations into account to detect failing measures automatically ([4]). By that strategy it is possible to get good registration results even with very bad initial guesses. An evaluation of the starting position showed almost constant result quality while increasing the initial error up to 43 mm. To our knowledge, such high attraction ranges have not yet been published in medical image registration. Furthermore, the algorithm needs no data-dependent modifications and solves the registration problem in a generic manner. This implies a high practical applicability for clinical applications. Future research topics will address the optimal choice of participating similarity measures to assure a democratic setup and the influence of history length for adapting the weights. A clinical evaluation will follow.

References

1. Livyatan, H., Yaniv, Z., Joskowicz, L.: Gradient-based 2-D/3-D rigid registration of flouroscopic X-ray to CT. IEEE Trans. on Med. Imaging 22(11), 1395–1405 (2003)
2. Viola, P.A.: Alignment by maximization of mutual information. International Journal of Computer Vision 24(2), 137–154 (2001)
3. Penney, G.P., Weese, J., Little, J.A., Desmedt, P., Hill, D.L.G., Hawkes, D.J.: A comparison of similarity measures for use in 2-D-3-D medical image registration. IEEE Trans. on Med. Imaging 17, 586–594 (1998)
4. Pluim, J., Maintz, J., Viergever, M.: Image registration by maximization of combined mutual information and gradient information. IEEE Trans. on Med. Imaging 19(8), 809–814 (2000)
5. Maintz, J., Viergever, M.: A survey of medical image registration. Medical Image Analysis 2(1), 1–36 (1998)
6. Garcia-Palomares, U.M., Gonzalez-Castaño, F.J., Burguillo-Rial, J.C.: A combined global & local search (CGLS) approach to global optimization. Journal of Global Optimization 34(3), 409–426 (2006)
7. Zheng, G., Zhang, X.: A unifying MAP-MRF framework for deriving new point similarity measures for intensity-based 2D-3D registration. In: The 18th International Conference on Pattern Recognition, vol. 2, pp. 1181–1185 (2006)
8. Feldmann, T., Bouattour, S., Paulus, D., Deinzer, F.: Kombination verschiedener Ähnlichkeitsmaße für die 2D/3D-Registrierung von Röntgenbildern mittels Demokratischer Integration. In: Bildverarbeitung für die Medizin. Informatik Aktuell, pp. 226–230. Springer, Heidelberg (2006)
9. Triesch, J., von der Malsburg, C.: Democratic integration: Self-organized integration of adaptive cues. Neural Computation 13(9), 2049–2074 (2001)
10. Jahn, J.: Vector optimization - theory, applications, and extensions. Springer, Heidelberg (2004)

Task-Optimal Registration Cost Functions

B.T. Thomas Yeo[1,⋆], Mert Sabuncu[1], Polina Golland[1], and Bruce Fischl[2,1]

[1] Computer Science and Artificial Intelligence Laboratory, MIT, USA
[2] Athinoula A. Martinos Center for Biomedical Imaging, MGH/HMS, USA
ythomas@csail.mit.edu

Abstract. In this paper, we propose a framework for learning the parameters of registration cost functions – such as the tradeoff between the regularization and image similiarity term – with respect to a specific task. Assuming the existence of labeled training data, we specialize the framework for the task of localizing hidden labels via image registration. We learn the parameters of the weighted sum of squared differences (wSSD) image similarity term that are optimal for the localization of Brodmann areas (BAs) in a new subject based on cortical geometry. We demonstrate state-of-the-art localization of V1, V2, BA44 and BA45.

1 Introduction

In medical imaging, registration is rarely the end-goal, and therefore the quality of image registration should be evaluated in the context of the application. The results of registration are usually used by other tasks, e.g., segmentation. Taking into account the parameters of the registration cost function has been shown to improve alignment as measured by the performance of subsequent tasks, such as population analysis [1] and segmentation [2,3]. This paper proposes a framework for optimizing parameters of registration cost functions for a specific task.

A common image similarity measure used in registration is the weighted sum of squared differences (wSSD). wSSD assumes an independent Gaussian distribution on the image noise with the weights corresponding to the reciprocal of the variance. The weights are typically set to a constant global value [4,5]. Alternatively, a spatially-varying variance can be estimated from the intensities of registered images [6]. However, the estimated variance depends on the wSSD-regularization tradeoff: weaker deformation regularization leads to better intensity alignment and lower variance estimates [3].

Recent work in probabilistic template construction resolves this problem by marginalizing the tradeoff under a Bayesian framework [7] or estimating the tradeoff with the Minimum Description Length principle [8]. Since these methods are not guided by any task, it is unclear whether the resulting parameters are optimal for any specific task. After all, the optimal parameters for segmentation might be different from those for group analysis. In contrast, [9] proposes a generative model for segmentation, so the registration parameters are Bayesian-optimal for segmentation. When considering a single global tradeoff parameter,

⋆ Corresponding author.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 598–606, 2009.

an exhaustive search by cross-validation of segmentation accuracy is possible [3]. Unfortunately, an exhaustive search is not feasible for multiple parameters.

Unlike these generative approaches, we take the discriminative approach of directly incorporating the task in the parameter estimation. We assume that the performance of a particular task can be measured by a cost function g given the output of registration. Our method learns the parameters of a given registration cost function f that yield better registration of a new image with respect to a specific task. The task-specific cost function g is evaluated with information from training data that is not available to the registration cost function f.

Our formulation is related to the computation of the entire space of solutions of learning problems (e.g. SVM) as a function of a single regularization parameter [10]. Because we deal with multiple parameters, it is impossible for us to compute a solution manifold. Instead, we trace a path within the solution manifold that improves the task-specific cost function.

In this paper, we propose a framework that optimizes parameters of registration cost functions for a specific task. We learn the weights of the wSSD registration cost function to optimize the prediction of Brodmann Areas (BAs) in a new subject, effectively estimating the tradeoff between the similarity measure and regularization. We demonstrate improvement over existing localization methods [11] for several BAs.

2 Task-Optimal Registration

Given an atlas coordinate frame and a new image, $f(w, \Psi)$ denotes a smooth registration cost function parametrized by the weights w and transformation Ψ. For example, the parameters w can be the tradeoff between the regularization and image similarity measure. f is typically a function of the template and the input image, but we omit this dependency to simplify notation. We assume a known and fixed template. Image registration is the process of estimating Ψ^* for a given set of parameters w:

$$\Psi^*(w) = \underset{\Psi}{\operatorname{argmin}}\, f(w, \Psi). \tag{1}$$

A different set of parameters w will result in a different solution and thus will effectively lead to a different image coordinate system. While there are typically multiple solutions to Eq. (1), we work with a single local optimum in practice.

The results of registration are used for further tasks, such as image segmentation. We assume the task performance can be measured by a smooth cost function g, so that smaller values of $g(\Psi^*(w))$ correspond to better task performance. g is a function of input data associated with a subject, such as its anatomical labels or functional activation map, not available to the cost function f.

Given a set of training subjects, we seek the parameters w^* that generalize well to a new subject, i.e., registration of a new subject with w^* yields the transformation $\Psi^*(w^*)$ with a small task-specific cost $g(\Psi^*(w^*))$:

$$w^* = \underset{w}{\operatorname{argmin}}\, G(w) \qquad \text{where} \qquad G(w) \triangleq \sum_{s=1}^{S} g_s(\Psi_s^*(w)) + \operatorname{Reg}(w). \tag{2}$$

$\Psi_s^*(w)$ denotes the registration of training subject s with fixed weights w. g_s is the evaluation of g for training subject s. $\mathrm{Reg}(w)$ denotes regularization on w.

2.1 Optimizing Registration Parameters w

In this section, we discuss the optimization of the cost function in Eq. (2). Let ∂_x, ∂_x^2, and $\partial_{x,y}^2$ denote partial derivatives and $\Psi^*(w_0)$ denote a local minimum of the registration cost function for a fixed $w = w_0$.

Suppose we perturb w_0 by δw. Let $\Psi^*(w_0) + \delta\Psi^*(w_0, \delta w)$ denote the new locally optimal deformation for the new parameters $w_0 + \delta w$. Because of numerous local optima, $\Psi^*(w_0) + \delta\Psi^*(w_0, \delta w)$ might be far from $\Psi^*(w_0)$. If the Hessian $\partial_\Psi^2 f(w_0, \Psi)$ is positive definite at $\Psi = \Psi^*(w_0)$, then by the Implicit Function Theorem [12], a unique function $\delta\Psi^*(w_0, \delta w)$ exists with the same order of smoothness as f for small enough values of $\|\delta w\|$, such that $\delta\Psi^*(w_0, 0) = 0$.

Consequently, at $(w_0, \Psi^*(w_0))$ with positive definite Hessian, one can compute the derivatives of $\delta\Psi^*$, allowing us to traverse a curve of local optima, finding values of w that improve the task-specific cost function for the training images. Since the derivatives at any local optimum is zero, we can show that

$$\partial_w \Psi^*\Big|_{w_0} = -\left(\partial_\Psi^2 f(w_0, \Psi)\Big|_{\Psi^*(w_0)}\right)^{-1} \partial_{w,\Psi}^2 f(w, \Psi)\Big|_{w_0, \Psi^*(w_0)}. \tag{3}$$

In practice, the matrix inversion in Eq. (3) is computationally prohibitive for high-dimensional warps Ψ. As a result, we consider a simplification of Eq. (3) by setting the Hessian to be the identity:

$$\partial_w \Psi^*\Big|_{w_0} \approx -\partial_{w,\Psi}^2 f(w, \Psi)\Big|_{w_0, \Psi^*(w_0)}. \tag{4}$$

Since $-\partial_\Psi f$ is the direction of gradient descent of the cost function Eq. (1), we can interpret Eq. (4) as approximating the new local minimum to be in the same direction as the *change* in the direction of gradient descent as w is perturbed. Differentiating the cost function in Eq. (2), using the chain rule, we get

$$\partial_w G = \partial_w \left(\sum_{s=1}^{S} g_s(\Psi_s^*(w)) + \mathrm{Reg}(w)\right) = \sum_{s=1}^{S} [\partial_{\Psi_s^*} g_s][\partial_w \Psi_s^*] + \partial_w \mathrm{Reg}(w). \tag{5}$$

We can therefore optimize Eq. (2) by standard gradient descent. We summarize the training procedure of the task-optimal image registration framework below:

- Initialize w to uniform values.
- Estimate $\Psi_s^*(w) = \mathrm{argmin}_{\Psi_s} f_s(w, \Psi_s)$, i.e., perform registration of each training subject s.
- Iterate until convergence:
  - Given current estimates $(w, \{\Psi_s^*(w)\})$, compute the gradient $\partial_w G$ in Eq. (5) using $\partial_w \Psi^*$ in Eq. (4).
  - Perform line search in the direction opposite to the gradient $\partial_w G$.

Each line search involves evaluating the cost function G multiple times, which in turn requires registering the training subjects. Since we are initializing from a local optimum, for a small change in w, the registration converges quickly.

Since nonlinear registration is dependent on initialization, the current estimates $(w, \Psi^*(w))$, which were initialized from previous estimates, might not be achievable when initializing the registration with the identity transform. The corresponding parameters w might therefore *not* generalize well to a new subject initialized with the identity transform. Consequently, after every few iterations, we re-initialize the transformations to the identity transform, re-register the images and check that G is better than the current best value of G computed with initialization from the identity transform.

Remark: Degeneracies can arise for local minima with a singular Hessian. For example, let $\Psi = [a\ b]$ and $f(\Psi, w) = (ab - w)^2$. Then the determinant of the Hessian at *any* local minimum is equal to zero! In this case, there is an infinite number of local minima near the current local minimum $\Psi^*(w)$, i.e., the gradient is not defined - our algorithm might then be stuck in the current estimates of w. In our experiments, these degeneracies do not seem to pose serious problems.

3 Learning wSSD for Hidden Label Alignment

We now instantiate the task-optimal registration framework for localizing hidden labels in images. Here, we work with meshes modeling the cortical surface, although it should be clear that the discussion extends to volumetric images. We assume that the meshes have been spherically parameterized and represented as spherical images: a geometric attribute is associated with each mesh vertex, describing local cortical geometry.

Suppose we have a set of spherical training images $\{I_s\}$ with a particular structure manually labeled. We represent the binary labels as signed distance transforms on the sphere $\{L_s\}$. We assume the existence of a spherical image template I_T and corresponding distance transform L_T. In this paper, we select one of the training subjects as the template. Our task is to align a new image to the template and predict the boundary of the hidden structure in the new subject by transferring the labels from the template to the new subject.

We represent the transformation Ψ as a composition of diffeomorphic warps, each parameterized by a stationary velocity field [5,13]. A diffeomorphic warp Φ is associated with a smooth stationary velocity field v via a stationary ODE: $\partial_t \Phi(x,t) = v(\Phi(x,t))$ with an initial condition $\Phi(x,0) = x$. The solution at $t = 1$ is denoted as $\Phi(x,1) = \Phi(x) = \exp(v)(x)$, where we have dropped the time index. A solution can be computed efficiently using scaling and squaring [14].

For a given image I_s, we define the registration cost function:

$$f_s(w, \Psi) = \sum_i w_i^2 \left[I(x_i) - I_s(\Psi(x_i))\right]^2 + \sum_i \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left(\frac{\|\Psi(x_i) - \Psi(x_j)\| - d_{ij}}{d_{ij}} \right)^2,$$

where $\Psi(x_i)$ denotes the point on the sphere S^2 to which Ψ maps the point $x_i \in S^2$. The first term corresponds to the wSSD image similiarity. The

parameterization of the weights as w_i^2 ensures non-negative weights. The second term is the regularization on the transformation Ψ. $\mathcal{N}_i$ is a predefined neighborhood around vertex i, d_{ij} is the original distance between the neighbors $\|x_i - x_j\|$. A higher weight w_i corresponds to placing more emphasis on matching the cortical geometry of the template at spatial location x_i relative to the regularization.

To register subject s to the template, let Ψ_0 be the current estimate of Ψ. We seek an update $\exp(v)$ parameterized by a stationary velocity field v:

$$f_s(w, \Psi_0 \circ \exp(v)) = \sum_i w_i^2 \left[I_T(x_i) - I_s(\Psi_0 \circ \exp(v)(x_i))\right]^2 + \sum_i \frac{1}{|\mathcal{N}_i|} \sum_{j \in \mathcal{N}_i} \left(\frac{\|\Psi_0 \circ \exp(v)(x_i) - \Psi_0 \circ \exp(v)(x_j)\| - d_{ij}}{d_{ij}} \right)^2. \tag{6}$$

We adopt the techniques in the Spherical Demons algorithm [13] to differentiate Eq. (6) with respect to v, evaluated at $v = 0$. Defining $\nabla I_s(\Psi_0(x_i))$ to be the gradient of the warped image $I_s(\Psi_0(\cdot))$ at x_i, $\nabla \Psi_0(x_i)$ to be the Jacobian matrix of Ψ_0 at x_i and v_i to be the velocity vector tangent to vertex x_i, we get

$$\partial_{v_i} f_s(w, \Psi_0 \circ \exp(v))\Big|_{v=0} = -2w_i^2 \left[I_T(x_i) - I_s(\Psi_0(x_i))\right] \left[\nabla I_s(\Psi_0(x_i))\right]^T + 2 \sum_{j \in \mathcal{N}_i} \left(\frac{1}{|\mathcal{N}_i|} + \frac{1}{|\mathcal{N}_j|} \right) \left(\frac{\|\Psi_0(x_i) - \Psi_0(x_j)\| - d_{ij}}{d_{ij}^2 \|\Psi_0(x_i) - \Psi_0(x_j)\|} \right) \left[\Psi_0(x_i) - \Psi_0(x_j)\right]^T \nabla \Psi_0(x_i). \tag{7}$$

Eq. (7) instantiates $\partial_\Psi f_s$ for this application. We can then perform gradient descent of the registration cost function f_s to obtain Ψ_s^*, which can be used to evaluate the task-specific cost function g_s. We adopt a simple label similarity measure for our task of localizing hidden labels:

$$g_s(\Psi^*) = \sum_i \left[L_T(x_i) - L_s(\Psi_s^*(x_i))\right]^2, \tag{8}$$

A low value of g_s indicates good alignment of the hidden label maps between the template and subject s, suggesting good prediction of the hidden label.

Here, we ignore the regularization $\mathrm{Reg}(w)$, but still achieve good results. One reason is that the re-registration after every few line searches helps to regularize against bad values of w. There is also implicit regularization in the framework: for example, w cannot become arbitrary large, since registration achieved with almost no regularization will lead to poor task performance.

Given the current estimates (w, Ψ_s^*), to update w using Eq. (5), we evaluate:

$$\partial_{\Psi^*} g_s = \partial_{v_i} g_s(\Psi_s^* \circ \exp(v))\Big|_{v=0} = -2 \left[L_T(x_i) - L_s(\Psi_s^*(x_i))\right] \left[\nabla L_s(\Psi_s^*(x_i))\right]^T \tag{9}$$

$$\partial_w \Psi_s^* \approx -\partial^2_{w_i, v_j} f_s(w, \Psi_s^* \circ \exp(v))\Big|_{v=0} = 4w_i \left[I(x_i) - I_s(\Psi_s^*(x_i))\right] \left[\nabla I_s(\Psi_s^*(x_i))\right]^T \delta(i,j).$$

4 Experiments

In this section, we demonstrate the utility of the task-optimal registration framework for localizing Brodmann Areas (BAs). We compare the framework with using uniform weights [4,5] and FreeSurfer [6].

BAs are cyto-architectonically defined parcellations of the cerebral cortex closely related to cortical function. We consider 10 human brains analyzed via postmortem histology [15]. Histologically defined BAs were sampled onto each hemispheric surface model and sampling errors were manually corrected. In this paper, we consider V1 and V2, which are well-predicted by local geometry and the Broca's areas: BA44 and BA45, which are not [11].

Even though each subject has multiple BAs, we focus on each structure independently. This allows us to interpret the weights in the wSSD in association with a particular label: a large weight at a particular location implies that the cortical geometry at that spatial location of the template is significant for localizing the label of interest.

4.1 Methods

Task-Optimal. We perform leave-one-out cross validation to predict BA location. For each BA and a test subject, we use one of the remaining 9 subjects as the template and the remaining 8 subjects for training. Once the weights are learned, we use them to register the test subject and predict the BA of the test subject by transferring the BA label from the template to the subject. We compute the symmetric mean Hausdorff distance between the boundary of the true BA and the predicted BA on the cortical surface of the test subject – smaller Hausdorff distance corresponds to better localization. There are 90 possibilities to select the test subject and the template. Here, we consider 20 of the 90 possibilities by selecting each of the 10 subjects to be a test subject twice (with a different randomly selected template), resulting in a total of 20 trials and 20 mean Hausdorff distances for each BA and for each hemisphere.

Uniform-weight. We repeat the process for the uniform-weight method using the same 20 pairs of subjects, where all the w_i's are manually set to a global fixed value w without training. We explore 12 different values of global weight w, chosen so that the deformations range from rigid to flexible warps. For each BA and each hemisphere, we pick the *best* value of w leading to the lowest mean Hausdorff distances. Because there is no cross-validation in picking the weights, the uniform-weight method is using an unrealistic version of the strategy proposed in [3].

FreeSurfer. Finally, we use FreeSurfer [6] to register the 10 *ex vivo* subjects to the FreeSurfer Buckner40 atlas, constructed from the MRI of 40 *in vivo* subjects. Once registered into this *in vivo* atlas space, for the same 20 pairs of subjects, we can use the BAs of one *ex vivo* subject to predict another *ex vivo* subject. We note that FreeSurfer also uses the wSSD cost function, but a more sophisticated regularization that penalizes both metric and areal distortion. For a particular tradeoff between the similarity measure and regularization, the Buckner40 template consists of the empirical mean and variance of the 40 *in vivo* subjects registered to template space. We use the reported FreeSurfer tradeoff parameters that were used to produce prior state-of-the-art BA alignment [11].

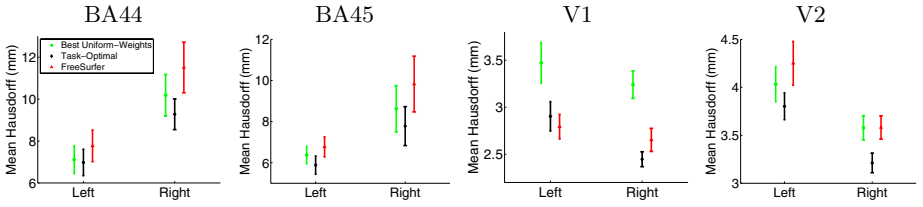


Fig. 1. Mean Hausdorff Distance (in mm) for V1, V2, BA44 and BA45. For the uniform-weight method, the result corresponding to the *best* weight is reported.

We run the task-optimal and uniform-weight methods on a low-resolution subdivided icosahedron mesh containing 2,562 vertices, whereas FreeSurfer results were computed on high-resolution meshes of more than 100k vertices. In our implementation, training on 8 subjects takes on average 4hrs on a standard PC (AMD Opteron, 2GHz, 4GB RAM). Despite the use of the low-resolution mesh, we achieve state-of-the-art localization accuracy.

4.2 Results

Fig. 1 shows the alignment errors for V1, V2, BA44 and BA45. Not surprisingly, we achieve better localization of BA44 and BA45 over the uniform-weight method and FreeSurfer, since local geometry poorly predicts the Broca's areas.

Since local cortical geometry is predictive of V1 and V2, we expect the three methods to perform similarly for V1 and V2. Surprisingly, we achieve improvement in V2 alignment over the uniform-weight method and FreeSurfer. Our method also significantly improves the alignment of V1 with respect to the uniform-weight method. Compared with FreeSurfer, we achieve slightly worse localization in the left hemisphere but better localization in the right. This suggests that even when local geometry is predictive of the hidden labels, so that anatomy-based registration is reasonable for localizing the labels, tuning the registration cost function can further improve the task performance.

Since our measure of localization accuracy uses the mean Hausdorff distance, ideally we should incorporate it into our task-specific objective function instead of the SSD on the distance transform representing the BA. Unfortunately, the resulting derivative is difficult to compute and the gradient will be zero everywhere except at the BA boundaries, resulting in a slow optimization.

We note that the task-optimal and uniform-weights registrations are pairwise, while FreeSurfer registrations are performed via an atlas. In our experience, registration via an unbiased atlas is usually more accurate than direct pairwise registration. Furthermore, FreeSurfer utilizes atlas-based registration by default, and this was used to produce prior best BA alignment [11]. We also note that our approach allows the computation of multiple task-optimal templates, thus complementing recent approaches of using multiple atlases for segmentation [16,17].

5 Conclusion

In this paper, we present a framework for optimizing the parameters of any differentiable family of registration cost functions with respect to a specific task. The only requirement is that the task performance can be encoded by a differentiable cost function. We demonstrate state-of-the-art Brodmann area localization by optimizing the weights of the wSSD image-similarity measure. Future work involves applying the framework to other Brodmann areas and fMRI-defined functional regions, as well as estimating the optimal template in addition to the weights of the registration cost function. We also hope to design task-specific cost functions for tasks other than segmentation.

Acknowledgments. We thank Hartmut Mohlberg, Katrin Amunts and Karl Zilles for providing the histological dataset. Support for this research is provided in part by the NAMIC (NIH NIBIB NAMIC U54-EB005149), the NAC (NIH NCRR NAC P41-RR13218), the mBIRN (NIH NCRR mBIRN U24-RR021382), the NIH NINDS R01-NS051826 grant, the NSF CAREER 0642971 grant, NCRR (P41-RR14075, R01 RR16594-01A1), the NIBIB (R01 EB001550, R01EB006758), the NINDS (R01 NS052585-01) and the MIND Institute. Additional support was provided by The Autism & Dyslexia Project funded by the Ellison Medical Foundation. B.T. Thomas Yeo is funded by the A*STAR, Singapore.

References

1. Makrogiannis, S., Verma, R., Karacali, B., Davatzikos, C.: A joint transformation and residual image descriptor for morphometric image analysis using an equivalence class formulation. In: Proc. MMBIA (2006)
2. Commowick, O., Stefanescu, R., Fillard, P., Arsigny, V., Ayache, N., Pennec, X., Malandain, G.: Incorporating statistical measures of anatomical variability in atlas-to-subject registration for conformal brain radiotherapy. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 927–934. Springer, Heidelberg (2005)
3. Yeo, B., Sabuncu, M., Desikan, R., Fischl, B., Golland, P.: Effects of Registration Regularization and Atlas Sharpness on Segmentation Accuracy. Medical Image Analysis 12(5), 603–615 (2008)
4. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased Diffeomorphic Atlas Construction for Computational Anatomy. NeuroImage 23, 151–160 (2004)
5. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1), S61–S72 (2009)
6. Fischl, B., Sereno, M., Tootell, R., Dale, A.: High-resolution Intersubject Averaging and a Coordinate System for the Cortical Surface. Human Brain Mapping 8(4), 272–284 (1999)
7. Allassonniere, S., Amit, Y., Trouvé, A.: Toward a Coherent Statistical Framework for Dense Deformable Template Estimation. Journal of the Royal Statistical Society, Series B 69(1), 3–29 (2007)
8. Twining, C., Cootes, T., Marsland, S., Petrovic, V., Schestowitz, R., Taylor, C.: A Unified Information-Theoretic Approach to Groupwise Non-rigid Registration and Model Building. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 1611–3349. Springer, Heidelberg (2005)

9. Van Leemput, K.: Probabilistic Brain Atlas Encoding Using Bayesian Inference. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 704–711. Springer, Heidelberg (2006)
10. Park, M., Hastie, T.: l_1-regularization Path Algorithm for Generalized Linear Models. J. of the Royal Stat. Soc. B 69, 659–677 (2007)
11. Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B., Mohlberg, H., Amunts, K., Zilles, K.: Cortical Folding Patterns and Predicting Cytoarchictecture. Cerebral Cortex 18(8), 1973–1980 (2008)
12. Rudin, W.: Principles of Mathematical Analysis (1976)
13. Yeo, B., Sabuncu, M., Vercauteren, T., Ayache, N., Fischl, B., Golland, P.: Spherical Demons: Fast Surface Registration. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 745–753. Springer, Heidelberg (2008)
14. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A Log-Euclidean Framework for Statistics on Diffeomorphisms. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 924–931. Springer, Heidelberg (2006)
15. Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H., Zilles, K.: Broca's Region Revisited: Cytoarchitecture and Intersubject Variability. Journal of Comparative Neurology 412(2), 319–341 (1999)
16. Heckemann, R., Hajnal, J., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain mri segmentation combining label propagation and decision fusion. NeuroImage 33(1), 115–126 (2006)
17. Sabuncu, M., Balci, S., Golland, P.: Discovering Modes of an Image Population through Mixture Modeling. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 381–389. Springer, Heidelberg (2008)

Hybrid Spline-Based Multimodal Registration Using Local Measures for Joint Entropy and Mutual Information

Andreas Biesdorf[1], Stefan Wörz[1], Hans-Jürgen Kaiser[2], Christoph Stippich[3], and Karl Rohr[1]

[1]University of Heidelberg, BIOQUANT, IPMB, and DKFZ Heidelberg
Dept. Bioinformatics and Functional Genomics, Biomedical Computer Vision Group
[2]University Hospital, RWTH Aachen University, Dept. of Nuclear Medicine
[3]University Hospital Heidelberg, Dept. of Neuroradiology

Abstract. We introduce a new hybrid approach for spline-based elastic registration of multimodal medical images. The approach uses point landmarks as well as intensity information based on local analytic measures for joint entropy and mutual information. The information-theoretic similarity measures are computationally efficient and can be optimized independently for each voxel. We have applied our approach to synthetic images, brain phantom images, as well as clinically relevant multimodal medical images. We also compared our measures with previous measures.

1 Introduction

Image registration is an important technique for medical diagnosis, surgical planning, and treatment control. A challenge is to cope with the broad range of applications as well as the large spectrum of imaging modalities. Previous approaches for image registration can be classified according to the transformation model (e.g., rigid or elastic) and the used image information (e.g., landmarks or intensities). *Landmark-based* approaches are computationally efficient and allow coping with large geometric differences, while *intensity-based* approaches use more image information and do not require segmentation. Concerning *intensity-based* approaches, it is important to distinguish between monomodal and multimodal registration problems, since different types of similarity metrics are required. *Monomodal* registration can be achieved by, e.g., using the sum of squared intensity differences (SSD). Registration of images of different modalities, however, requires *multimodal* similarity metrics such as joint entropy (JE) or mutual information (MI). In general, registration using such *information-theoretic* intensity similarity metrics is robust and accurate, however, these metrics are computationally expensive since they require the estimation of probability density functions based on joint histograms (e.g., [1,2,3]).

In recent years, increased attention has been paid to *hybrid* registration approaches that integrate both landmark and intensity information (e.g., [4,5,6,7,8,9,10,11]). Hybrid approaches can be classified into two classes. The first

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 607–615, 2009.

class of approaches uses the two types of information subsequently. Typically, one type of information is used to perform an affine preregistration to decrease computation time in the following elastic registration step (e.g., [4,5,6,7,8]). The second class of hybrid approaches couples both types of information directly (e.g., [9,10,11]). However, only few hybrid approaches have been designed for *multimodal* images (e.g., [9,5,7]) and, even if information-theoretic similarity measures are employed, often only application to monomodal images is reported. In addition, previous *spline-based* hybrid approaches typically use coarse physical deformation models such as B-splines (e.g., [9,7]) or thin-plate splines (e.g., [5,8]), and incorporate intensity information using *global*, computationally expensive information-theoretic measures. *Local* information theoretic measures exists (e.g., [12,13]), however, they have not been used in hybrid registration schemes.

In this contribution, we present a new *hybrid* approach for *spline-based* elastic registration of *multimodal* medical images. Our approach is formulated as an energy-minimizing functional that incorporates point landmarks and intensity information as well as a regularization using physically-based splines. The intensity information is evaluated *locally* based on analytic measures for joint entropy and mutual information. We suggest new information-theoretic measures that are modifications of the measures in [13], but which are more robust w.r.t. noise as well as more similar to the original information-theoretic formulations. In contrast to [13], where the measures are evaluated globally, we here evaluate and optimize the measures *locally*, i.e., independently for each voxel, and integrate them within a *hybrid* spline-based registration scheme.

2 Multimodal Hybrid Elastic Registration

2.1 Hybrid Energy Minimizing Functional

Our hybrid approach for elastic registration of *multimodal* images is based on an energy minimizing functional J_{Hybrid} that incorporates both landmark and intensity information and a regularization term:

$$J_{Hybrid}(\mathbf{u}) = J_{Data,I}(g_1, g_2, \mathbf{u}^I) + \lambda_I J_I(\mathbf{u}, \mathbf{u}^I) + \lambda_L J_{Data,L}(\mathbf{u}) + \lambda_E J_{El}(\mathbf{u}) \quad (1)$$

The first term $J_{Data,I}$ describes the intensity-based similarity measure between the source and target image, g_1 and g_2, respectively. With the second term J_I the intensity-based deformation field $\mathbf{u}^I$ is coupled with the final deformation field $\mathbf{u}$ using a weighted Euclidean distance. The term $J_{Data,L}$ incorporates the landmark information based on approximating Gaussian Elastic Body Splines (GEBS). In contrast to other splines, an advantage of GEBS is that cross-effects can be handled [14], i.e., transversal contractions lead to longitudinal dilations according to properties of elastic tissue. The fourth term J_{El} represents the regularization of the deformation field according to the Navier equation of linear elasticity, which constrains the deformation field to physically realistic deformations. The overall functional J_{Hybrid} is minimized alternatingly w.r.t. $\mathbf{u}^I$ and $\mathbf{u}$.

In previous work [11], $J_{Data,I}$ was defined based on the sum-of-squared intensity differences, thus the approach was only applicable to *monomodal* registration. Here, we present two new formulations of $J_{Data,I}$ for *multimodal* registration as well as new schemes for minimizing J_{Hybrid} w.r.t. $\mathbf{u}^I$.

2.2 Multimodal Registration Using Analytic Measures for Joint Entropy and Mutual Information

For the intensity-based similarity measure $J_{Data,I}$ in (1) we suggest to use *local* analytic measures for joint entropy and mutual information. Let g be an image of dimension d over the continuous domain $\Omega \subset \mathbb{R}^d$, and $N_R(\mathbf{x})$ be the neighborhood of radius R around a point $\mathbf{x} \in \Omega$. For R being sufficiently small, the first order Taylor approximation $T_g(\mathbf{x}) \simeq \nabla g(\mathbf{x})^T \cdot \mathbf{x} + g_0(\mathbf{x})$ can be used as an approximation of $g(\mathbf{x})$ in the neighborhood of $\mathbf{x}$, where ∇g denotes the image gradient. The intensities $g(\mathbf{x})$ in the neighborhood $N_R(\mathbf{x})$ can be characterized by a random variable $\mathbf{g}$, which can be described by a probability density function. In [13] it was shown that for two images g_1 and g_2 the joint entropy $JE(\mathbf{x})$ and the mutual information $MI(\mathbf{x})$ can then be approximated by

$$JE_{orig}(\mathbf{x}) : \ \{c_d + \log_2(\|\nabla g_1(\mathbf{x})\| \|\nabla g_2(\mathbf{x})\| |\sin(\theta)|)\} \rightarrow \min, \tag{2}$$

$$MI_{orig}(\mathbf{x}) : \ \{c_d - \log_2(|\sin\theta|)\} \rightarrow \max, \tag{3}$$

where θ represents the angle between $\nabla g_1(\mathbf{x})$ and $\nabla g_2(\mathbf{x})$, c_d is a constant that depends on the dimension d, and $\|\cdot\|$ denotes the Euclidean norm. However, due to properties of the logarithm, $JE_{orig}(\mathbf{x})$ and $MI_{orig}(\mathbf{x})$ yield very large absolute values if the argument $\|\nabla g_1(\mathbf{x})\| \|\nabla g_2(\mathbf{x})\| |\sin(\theta)|$ or $|\sin\theta|$ is close to zero which leads to unstable results. In [13], $JE_{orig}(\mathbf{x})$ and $MI_{orig}(\mathbf{x})$ were evaluated *globally*, i.e., the similarity for each voxel is integrated over the whole image. If, however, at a *single* voxel the argument of the logarithm is zero or close to zero, the global measure yields a very large value, since adding logarithmic terms is equivalent to the logarithm of the product of the arguments ($\sum_i \log_2 s_i = \log_2 \prod_i s_i$). Due to the singularity of the logarithm, the value at a single voxel can render the result of the metric useless. To circumvent this problem, in [13] simplified metrics have been used which omit the logarithm:

$$JE_{simp}(\mathbf{x}) : \left\{ \left(\nabla g_1^T(\mathbf{x}) \nabla g_2(\mathbf{x})\right)^2 \right\} \rightarrow \max \tag{4}$$

$$MI_{simp}(\mathbf{x}) : \left\{\cos^2\theta\right\} \rightarrow \max \tag{5}$$

A disadvantage of these measures is that they differ largely from the original measures in (2) and (3). Instead, we here propose alternative formulations for *JE* and *MI* that do not have a singularity and that are more similar to the original measures.

Joint Entropy Measure. For *JE*, we suggest an alternative measure to (2), where the logarithm is still included. To illustrate and motivate this modification, we have generated synthetic images simulating a multimodal registration

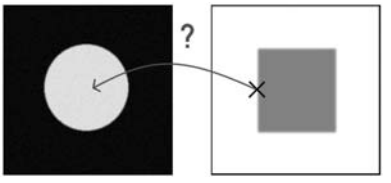


Fig. 1. Registration problem: Source and target images

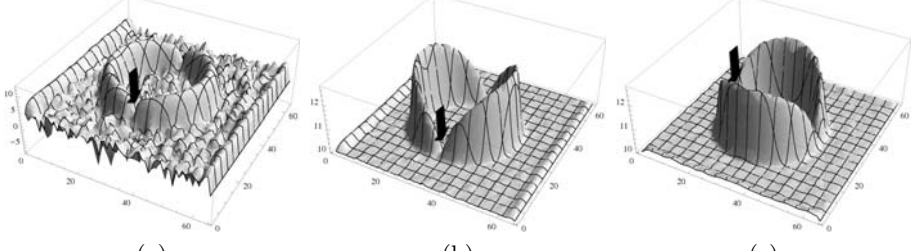
problem (light circle on a dark background and dark square on a light background, Fig. 1). By registration, the circle has to be stretched to the size and shape of the square. For each point in the source image, we have computed the similarity values w.r.t. the marked reference point (black cross) in the target image and have visualized them as 3D plots in Fig. 2. In each of the plots, the *intended* optimum is marked with a black arrow and circle. In (a), the result for $JE_{orig}(\mathbf{x})$ is shown. Since $JE_{orig}(\mathbf{x})$ needs to be minimized, the optimum can be found at the saddle point between the two elevated semi-circles. Obviously, this minimum is difficult to identify since it has a similar value as the background, which is very noisy. To reduce the sensitivity to noise, we introduce a constant $\epsilon > 0$, which is added to the argument. This leads to the result in (b). Now, the background noise is reduced significantly, but the optimum is still located at the saddle point between the two semi-circles, which is difficult to identify. To facilitate optimization, we replace $|\sin(\theta)|$ in (2) with $|\cos(\theta)|$, which leads to:

$$JE_{new}(\mathbf{x}) : \{\log_2 (\epsilon + \|\nabla g_1(\mathbf{x})\| \|\nabla g_2(\mathbf{x})\| |\cos(\theta)|)\} \rightarrow \max \tag{6}$$

Now, the measure has to be maximized, which is much easier (see (c)) than finding the saddle point in (b).

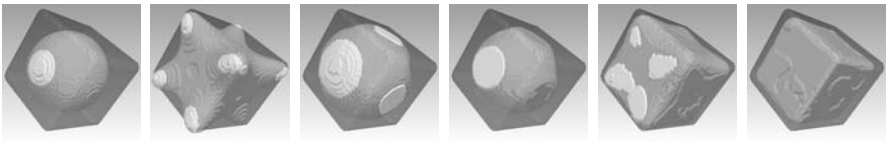
Mutual Information Measure. For *MI*, we suggest a similar modification where the logarithm is still included and a constant $\epsilon > 0$ is introduced. Here, the introduction of ϵ also improves the robustness of the metric:

$$MI_{new}(\mathbf{x}) : \{\log_2 |\epsilon + \sin\theta|\} \rightarrow \min \tag{7}$$



(a) (b) (c)

Fig. 2. Results for different metrics for *JE* for registration of the images in Fig. 1: (a) $JE_{orig}(\mathbf{x})$, (b) $JE_{orig}(\mathbf{x})$ after introducing a constant ϵ, (c) using $JE_{new}(\mathbf{x})$



(a) (b) (c) (d) (e) (f)

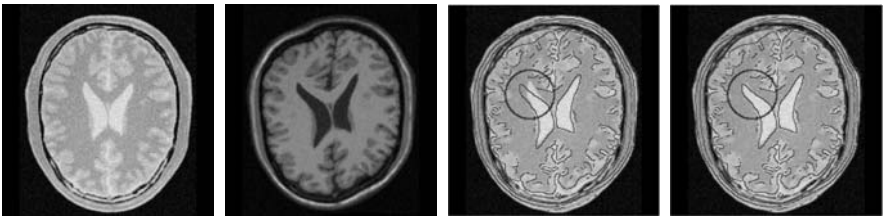
Fig. 3. Registration of a sphere with a cube. (a) Original source and target images, (b) after landmark-based registration, (c) after intensity-based registration using JE_{new}, (d) after intensity-based registration using MI_{new}, (e) after hybrid registration using JE_{new}, and (f) after hybrid registration using MI_{new}.

In our registration approach, the measures are integrated as $J_{Data,I}$ in (1). Optimization of J_{Hybrid} is performed alternatingly w.r.t. $\mathbf{u}^I$ and $\mathbf{u}$. For minimization w.r.t. $\mathbf{u}^I$, $J_{Data,I} + \lambda_I J_I$ has to be minimized. To this end we have derived analytic expressions for the partial derivatives of the different measures. Note that in contrast to [13], in our approach MI and JE are evaluated *locally*, i.e., optimization of $\mathbf{u}^I$ is computed independently for each voxel, which improves the efficiency. In comparison to [11], we here use a minimization scheme based on steepest gradient since the analytic information-theoretic measures cannot be formulated by squared differences and thus the method of Levenberg-Marquardt cannot be used.

3 Experimental Results

We have applied our registration approach to 3D synthetic images, 3D brain phantom images, and clinically relevant 3D medical images. In a first experiment, we have applied our approach to the 3D synthetic sphere and cube shown in Fig. 3a. The images have a size of $128 \times 128 \times 128$ voxels and inverted contrast to simulate a multimodal registration problem. Thus, a monomodal similarity measure (e.g., SSD) would fail. Note that registration is difficult since the volume of the cube is nearly two times larger than the volume of the sphere. Eight landmarks were defined at the corners of the cube, and the registration result is computed based on the different metrics. Fig. 3 shows the results for landmark-based registration (b), for intensity-based registration based on JE_{new} (c) and MI_{new} (d), as well as for hybrid registration based on JE_{new} (e) and MI_{new} (f). It can be seen that JE_{new} and MI_{new} yield similar results. For landmark-based registration, only the corners of the cube are aligned. Using intensity-based registration, the faces of the sphere are aligned, but not the corners. The best result is obtained using the new hybrid approach, since both the faces as well as the corners are aligned.

To compare the results using our new measures with those in [13], we evaluated the registration accuracy for intensity-based registration of multimodal phantom images (MRI-PD and MRI-T1) from the BrainWeb database [15] for different levels of Gaussian noise ($\sigma_n = 0, 5, 9$; see Fig. 4a,b for the case of $\sigma_n = 5$). Here, we present results for the comparison of MI_{new} and MI_{simp}. We have generated



(a) (b) (c) (d)

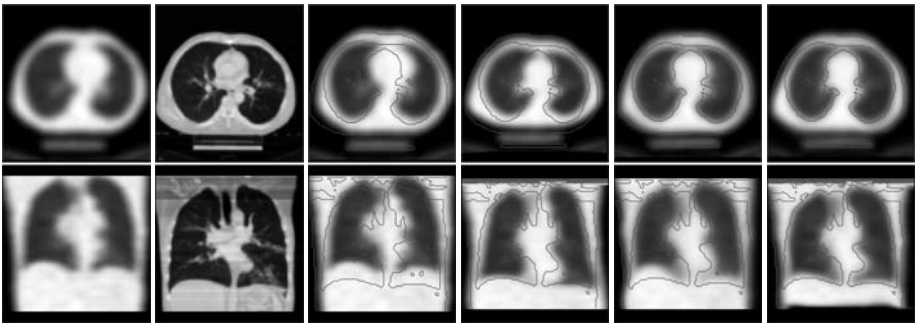
Fig. 4. Comparison of MI_{simp} and MI_{new}: (a) Source and (b) target image for $\sigma_n = 5$. Registration result and edge overlay of the target image for (c) MI_{simp} and (d) MI_{new}.

a deformation field $\mathbf{u}_o$ based on eight landmarks using GEBS, which is applied to the different MRI-T1 images to obtain three target images with known elastic deformations and different levels of noise. After registration, we compared the computed deformation $\mathbf{u}$ with the original deformation $\mathbf{u}_o$ and quantified the registration accuracy by the mean geometric error $\overline{e}_g = \overline{\|\mathbf{u}_o - \mathbf{u}\|}$. Note that in this experiment we did not use landmarks. Without registration, we have $\overline{e}_g = 3.40$ pixels. For $\sigma_n = 0$, registration using MI_{simp} and MI_{new} yields $\overline{e}_g = 1.44$ and $\overline{e}_g = 1.08$ pixels, respectively. With increasing level of noise, our new measure consistently yields more accurate results than the previous measure. Fig. 4, for example, shows the registration results for $\sigma_n = 5$ for MI_{simp} (c) and MI_{new} (d). It can be seen that registration is more accurate in the neighborhood of the ventricles when using MI_{new}. The visual result agrees with the quantitative values: $\overline{e}_g = 1.85$ for MI_{simp}, and $\overline{e}_g = 1.55$ for MI_{new}. For $\sigma_n = 9$, MI_{simp} yields $\overline{e}_g = 2.30$, while MI_{new} yields $\overline{e}_g = 1.78$. Thus, our new measure MI_{new} is more robust w.r.t. noise than MI_{simp}. We also investigated the effect of ϵ on the registration result of MI_{new}. We found that the choice of ϵ is not critical, however, best results were obtained for $0.5 \leq \epsilon \leq 1.5$.

Furthermore, we quantified the registration accuracy of our approach using real, clinically relevant 3D multimodal medical images. We used five pairs of 3D PET transmission ($170 \times 128 \times 128$ voxels) and 3D CT ($320 \times 400 \times 400$ voxels) images of the human chest. Fig. 5, for example, shows two orthogonal slices of a 3D PET (a) and 3D CT (b) image before registration, as well as overlay images of the PET image with computed edges of the CT image before (c), after landmark-based (d), after intensity-based (e), and after hybrid registration (f) using MI_{new}. It can be seen that both the landmark-based and intensity-based registration schemes improve the alignment significantly, however, the best result

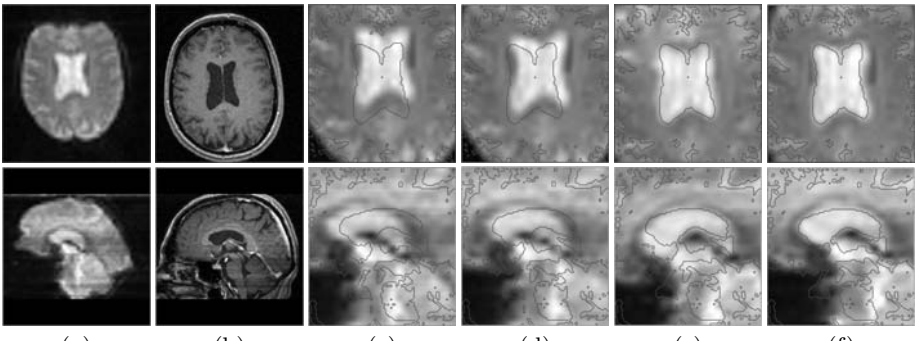
Table 1. Registration of 3D PET and CT images: Mean geometric error $\overline{e}_g$ (in voxel) for different metrics using different types of image information

			Intensities				**Hybrid**			
	Unreg.	**LM**	JE_{simp}	JE_{new}	MI_{simp}	MI_{new}	JE_{simp}	JE_{new}	MI_{simp}	MI_{new}
$\overline{e}_g$	7.36	5.32	5.60	5.76	6.70	5.60	5.17	5.05	5.28	4.83



(a) (b) (c) (d) (e) (f)

Fig. 5. Orthogonal slices of original (a) 3D PET transmission and (b) CT images. Source image with edge overlay of the target image (c) before, (d) after landmark-based, (e) after intensity-based, and (f) after hybrid registration using MI_{new}



(a) (b) (c) (d) (e) (f)

Fig. 6. Transversal and saggital slices of original (a) 3D fMRI and (b) MRI-T1 images. Section of the source image with edge overlay of the target image (c) before, (d) after landmark-based, (e) after intensity-based, and after (f) hybrid registration using MI_{new}.

is obtained when using the new hybrid scheme. This was also confirmed quantitatively. For this purpose, we identified ten pairs of corresponding anatomical point landmarks in each source and target image and computed the mean geometric registration error $\overline{e}_g$ (see Tab. 1). Before registration, we have $\overline{e}_g = 7.36$ voxels. Using landmarks only, we have $\overline{e}_g = 5.32$ voxels, which is a relative improvement of 27.7%. Using intensity information only, we obtain $\overline{e}_g = 5.60$ (23.8%) and $\overline{e}_g = 5.76$ (21.7%) voxels for JE_{simp} and JE_{new}, respectively. For MI_{simp} and MI_{new} we obtain $\overline{e}_g = 6.70$ (9.0%) and $\overline{e}_g = 5.60$ (24.0%) voxels. When using the hybrid approach, we obtain $\overline{e}_g = 5.17$ (29.7%) for JE_{simp} and $\overline{e}_g = 5.05$ (31.4%) for JE_{new} as well as $\overline{e}_g = 5.28$ (28.2%) for MI_{simp} and $\overline{e}_g = 4.83$ (34.3%) for MI_{new}. Thus, in all cases, the hybrid approach is more accurate than the intensity-based and the landmark-based approaches. Furthermore, the

new intensity similarity measures JE_{new} and MI_{new} yield more accurate results for the hybrid approach than the previous measures JE_{simp} and MI_{simp}. Overall, MI_{new} yields the best results.

We also applied our approach to register 3D MRI-T1 and functional MRI (fMRI) images ($144 \times 256 \times 256$ voxels). Fig. 6 shows a transversal and saggital view of 3D fMRI (a) and MRI-T1 (b) images. Before registration (c), a significant misalignment between source and target image is visible. After landmark-based registration using seven landmarks, a slight improvement is visible in the neighborhood of the ventricles (d). After intensity-based registration using MI_{new} (e), the global alignment is improved significantly, although local misalignments still exist. The best result is obtained when using our new hybrid approach (f), since both the global as well as the local alignment is improved.

4 Discussion

We introduced a new hybrid spline-based approach for elastic registration of multimodal images. The approach incorporates point landmarks, intensity information, as well as a physically-based regularization. Since the approach uses new local analytic measures for mutual information and joint entropy, the required derivatives can be calculated analytically, and optimization can be performed independently for each voxel. We have demonstrated the applicability of our approach using 3D synthetic images, 3D phantom images, and different types of real 3D medical images. It turned out that our new similarity measures yield a more accurate registration result than previously proposed measures. We also found that the hybrid approach improves the results compared to a pure intensity-based and a pure landmark-based scheme.

Acknowledgments. Support of the Deutsche Forschungsgemeinschaft (DFG) within the project ELASTIR (RO 2471/2) is gratefully acknowledged.

References

1. Viola, P., Wells, W.M.: Alignment by maximization of mutual information. In: Proc. ICCV 1995, Boston, MA, USA, pp. 16–23. IEEE Computer Society Press, Los Alamitos (1995)
2. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imaging 16(2), 187–198 (1997)
3. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: application to breast mr images. IEEE Trans. Med. Imaging 18(8), 712–721 (1999)
4. Johnson, H., Christensen, G.: Consistent landmark and intensity-based image registration. IEEE Trans. Med. Imag. 21(5), 450–461 (2002)
5. Wang, X., Feng, D.D.: Automatic Hybrid Registration for 2-D CT Abdominal Images. In: Proc. ICIG 2004, Hong Kong, China, pp. 208–211. IEEE Computer Society Press, Los Alamitos (2004)

6. Rohr, K., Cathier, P., Wörz, S.: Elastic registration of electrophoresis images using intensity information and point landmarks. Pattern Recognition 37(5), 1035–1048 (2004)
7. Teng, C.-C., Shapiro, L.G., Kalet, I.: Head and neck lymph node region delineation using a hybrid image registration method. In: Proc. ISBI 2006, Arlington, VA, USA, pp. 462–465 (2006)
8. Azar, A., Xu, C., Pennec, X., Ayache, N.: An interactive hybrid non-rigid registration framework for 3d medical images. In: Proc. ISBI 2006, Arlington, VA, USA, pp. 824–827 (2006)
9. Hartkens, T., Hill, D., Castellano-Smith, A., Hawkes, D., Maurer, C., Martin, A., Hall, W., Liu, H., Truwit, C.: Using points and surfaces to improve voxel-based non-rigid registration. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2489, pp. 565–572. Springer, Heidelberg (2002)
10. Kybic, J., Unser, M.: Fast parametric elastic image registration. IEEE Trans Image Processing 12(11), 1427–1442 (2003)
11. Wörz, S., Rohr, K.: Hybrid Physics-Based Elastic Image Registration Using Approximating Splines. In: Proc. MI 2008: Image Processing, San Diego, CA/USA (2008)
12. Hermosillo, G., Chef d'Hotel, C., Faugeras, O.D.: Variational methods for multimodal image matching. Internat. J. of Computer Vision 50(3), 329–343 (2002)
13. Karaçali, B.: Information theoretic deformable registration using local image information. Internat. J. of Computer Vision 72, 219–237 (2007)
14. Wörz, S., Rohr, K.: Physics-based elastic registration using non-radial basis functions and including landmark localization uncertainties. Computer Vision and Image Understanding 111, 263–274 (2008)
15. Kwan, R.K.-S., Evans, A.C., Pike, G.B.: MRI simulation-based evaluation of image-processing and classification methods. IEEE Trans. Med. Imaging 18(11), 1085–1097 (1999)

A Robust Solution to Multi-modal Image Registration by Combining Mutual Information with Multi-scale Derivatives

Philip A. Legg[1], Paul L. Rosin[1], David Marshall[1], and James E. Morgan[2]

[1] School of Computer Science, Cardiff University, UK
P.A.Legg@cs.cardiff.ac.uk, Paul.Rosin@cs.cardiff.ac.uk,
Dave.Marshall@cs.cardiff.ac.uk
[2] School of Optometry and Vision Sciences, Cardiff University, UK
morganje3@cardiff.ac.uk

Abstract. In this paper we present a novel method for performing image registration of different modalities. Mutual Information (MI) is an established method for performing such registration. However, it is recognised that standard MI is not without some problems, in particular it does not utilise spatial information within the images. Various modifications have been proposed to resolve this, however these only offer slight improvement to the accuracy of registration. We present Feature Neighbourhood Mutual Information (FNMI) that combines both image structure and spatial neighbourhood information which is efficiently incorporated into Mutual Information by approximating the joint distribution with a covariance matrix (c.f. Russakoff's Regional Mutual Information). Results show that our approach offers a very high level of accuracy that improves greatly on previous methods. In comparison to Regional MI, our method also improves runtime for more demanding registration problems where a higher neighbourhood radius is required. We demonstrate our method using retinal fundus photographs and scanning laser ophthalmoscopy images, two modalities that have received little attention in registration literature. Registration of these images would improve accuracy when performing demarcation of the optic nerve head for detecting such diseases as glaucoma.

1 Introduction

Image Registration is the task of finding a spatial transformation that aligns two (or more) images into matching correspondence. Within the medical domain, combining images of different modalities can provide a clinician with greater insight when treating a patient. Mutual Information (MI) is a widely recognised technique for registering different modalities, based on the entropy of the image regions being compared. Simultaneously proposed by Viola [1] and Collignon [2], MI relies on a statistical comparison between the images rather than individual intensity comparison, making it a suitable similarity measure for multi-modal images. Given our floating image A, and the region being covered by this in

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 616–623, 2009.

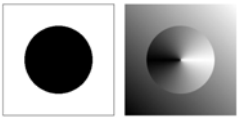


Fig. 1. Two images with clear structural relationship but no simple intensity mapping

our reference image B, MI can then be defined as $I(A;B) = H(A) + H(B) - H(A,B)$, where $H(A)$ is the entropy of image A, $H(B)$ is the entropy of image B and $H(A,B)$ is the joint entropy of the two. We wish to find the spatial transformation that maximises $I(A;B)$.

Since being introduced, there have been many methods proposed to improve the registration accuracy of MI. This is largely to resolve the fact that no spatial information is considered within the measure meaning each pixel is considered independently of its neighbours. Given two multi-modal images, there may be little or no consistent intensity mapping that relates the two together due to the differences in lighting and acquisition (Fig. 1) [3]. Therefore it is sensible to also include structural information since intensity alone may not provide adequate information.

There have been many proposed methods that attempt to improve MI. Pluim suggests integrating a gradient value into the MI measure [4], however this is simply done by computing standard MI and multiplying this by a gradient term. Dowson incorporates interpolated pixel information to improve the entropy estimate [5] using NP-Windows. However this is computationally expensive giving a lengthy runtime that is unsuitable for our task. Rueckert proposes higher-order MI that computes entropy for intensity pairs rather than individual intensities, to introduce neighbouring pixel information into MI [6]. Beijing also uses higher-order MI [7] to include a range of other properties such as mean and median values of a pixel neighbourhood, different neighbouring pixels (left and right neighbours) and also a gradient value. Kubecka suggests using gradient-image MI, where MI is computed for both the original images (after performing illumination correction) and also for the corresponding gradient images [8].

Another proposed scheme to improve Mutual Information is Regional Mutual Information (RMI), introduced by Russakoff [9]. This takes neighbouring pixels into account to incorporate spatial information. Essentially, for each pixel, a vector of all the local intensities is created for both of the images being registered. While this provides much greater relational information for each pixel, it also means that many intensities need to be considered (e.g. for a neighbourhood radius of 1, this would mean 18 intensities per pixel when registering two 2-dimensional images). Using a joint histogram to represent this would be unfeasible and place far too great a demand on the physical memory of the computer system (using 8-bit images, the joint histogram would consist of $256^{18} \approx 2.2301 \times 10^{43}$ bins). To overcome this complexity problem, the samples are replaced by a covariance matrix which substantially reduces the amount of data. The covariance matrix represents the relation between the original vector elements by approximating the joint intensities by a normal distribution. This

method offers a clear improvement over standard MI. Yang [10] extends this method by using the same approach as Russakoff, but incorporating a mean neighbourhood intensity value instead of all intensities to reduce the complexity of the original matrix computation. However, this method fails to perform as well as RMI, clearly showing that too much information is lost by simplifying the data in this fashion.

In this paper, we incorporate multi-scale feature derivatives along with spatial neighbourhood knowledge into a MI framework. In doing this, we can encapsulate greater detail from our images in a much more efficient approach. We demonstrate our method using retinal fundus photographs and scanning laser ophthalmoscope (SLO) images, and obtain very high accuracy of registration. Successful registration would help improve demarcation and analysis of the optic nerve head, an important task in detecting such diseases as glaucoma.

2 Multi-scale Feature Neighbourhood MI

Feature Neighbourhood Mutual Information (FNMI) consists of firstly computing the features from our images and then combining these together in a Mutual Information framework. To be a successful similarity measure, the result should be maximised at the correct transformation. If we imagine the surface given by calculating the similarity measure at different transformation points, ideally this should also be smooth with few local maxima and good convergence to the maximum solution, so this can be efficiently found by an optimised search algorithm.

For our features, we compute the gradient magnitude of each image at multiple scales to detect structural change within the image. Typically, an image will consist of many different features, varying in shape and size. In the case of retinal images, we expect to see the optic nerve head along with many blood vessels, all of various thickness. In order to highlight all these features effectively, we incorporate the notion of scale-space [11]. Scale-space applies Gaussian blurring to an image at multiple kernel sizes (defined as σ) that can determine how features in the image relate at a larger scale. At a low scale we can pick out fine detail of small blood vessels and the edges of larger vessels. For larger vessels, there maybe little connectivity information that would link one edge with the other, so by using a larger scale smaller features become grouped giving a more generalized edge map showing the connectivity of the larger features.

2.1 Incorporating Features with Mutual Information

Our method for incorporating features is an adaptation of RMI [9]. We create a stack for each of the images being registered, consisting of the original image and its corresponding feature images (Fig. 2). It is possible to combine any number of feature images, where more features may improve accuracy further, although as more features are added this will affect computation time. Given the floating image and the area being registered to from the reference image, we create a

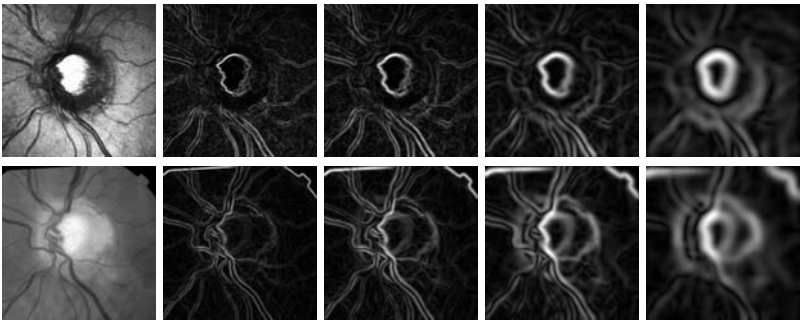


Fig. 2. Top: SLO image with multi-scale gradient ($\sigma = 1, 2, 4$ and 8). Bottom: Extract from fundus photograph with multi-scale gradient ($\sigma = 1, 2, 4$ and 8).

vector that consists of each pixel and its neighbours for each image in the stack. If we use f feature images, then the vector will consist of $d = 18 \times (f+1)$ elements (the pixel and its 8 neighbouring pixels for both the floating and reference images (18 points), for each feature image plus the original image). This is done for every point where the floating image and reference image overlap (ignoring background in the fundus photograph), and combined to give matrix P.

We subtract the mean from each point in the matrix, and calculate the covariance of the matrix, given by $C = \frac{1}{N} PP^T$. From [12], the entropy of a normally distributed set of points in $\Re^d$ with covariance matrix c is given by $H(c) = \log((2\pi)^{\frac{d}{2}} \det(c)^{\frac{1}{2}})$. The joint entropy is computed by $H(C)$, and the marginal entropies are computed by $H(C_A)$ and $H(C_B)$, where C_A is the $\frac{d}{2} \times \frac{d}{2}$ sub-matrix in the top-left corner of C, and C_B is the $\frac{d}{2} \times \frac{d}{2}$ sub-matrix in the bottom-right corner of C. Mutual Information is computed by $MI = H(C_A) + H(C_B) - H(C)$.

2.2 Registration Search

Searching the transformation space in registration tasks is difficult since the space is very large when considering all possible translations and rotations, meaning an exhaustive search is unfeasible. Instead, we adopt the Nelder-Mead simplex algorithm [13] to find translation. We know the rotation of the image to be within the range $\pm 3°$, which we step through at intervals of $0.5°$.

We incorporate a multi-resolution image pyramid to search for the correct registration on a coarse-to-fine approach. The coarse level is initialise by placing the floating image in the centre of the reference image since it is expected that the optic nerve head will occur near the centre of the fundus photograph. At the coarse level, we can obtain a result much faster due to a much smaller search space, and so can afford to search all possible rotations within range. The result is then used to initialise the next level down in the pyramid, giving good approximation to the maximum at the fine level. We also reduce the rotation range to $\pm 1°$ as we step down the pyramid to reduce unnecessary computation. This approach improves computation time at the fine level and also helps to

avoid local maxima points that a search algorithm could easily become trapped within, due to the reduced search space.

3 Testing and Results

For our testing, we have 135 image pairs that are to be registered. The image modalities are retinal fundus photographs and scanning laser ophthalmoscopy (SLO) images. The size of each fundus image is 564×367 and the size of each SLO image is 286×278. For each of the methods tested, we compare the mean and median registration errors between the obtained result and the ground truth values. The ground truth registrations were performed by hand and approved by an expert clinician. Since the mean can be affected by outliers and the median is insensitive to half of the data, we also use a 5-point grading scheme to assess each method, with an expert clinician classifying each registration as either 'excellent', 'very good', 'good', 'weak' or 'fail'.

The results shown in Table 1 indicate that our method can obtain a very high degree of accuracy for registration. It can be seen that many of the other methods have a high variability in the results, shown by a high mean and a low median result. While there are some successful registrations found with each method, the results show that other approaches are not as consistent as our proposed method. However, it is clear to note that our method requires much greater computation time to achieve such accuracy. This is due to generating the additional feature image as part of the registration process, a task which could be sped up by utilising the GPU.

In our testing, we use one feature image at scale $\sigma = 2$. We propose that our registration can be improved by incorporating additional features to our registration, such as two images at scale $\sigma = 2$ and $\sigma = 4$. Likewise, it is noted that RMI is computed using a neighbourhood radius of 1. This neighbourhood radius could be extended, which should improve on the original result. Table 2 shows our extended testing of our method with additional features and RMI using a larger radius. It can be seen that the mean error is reduced when using additional features, or when increasing the radius of RMI. However it is difficult

Table 1. Registration error (Translation (pixels), Rotation (degrees) and Runtime (seconds))

Method	**Mean**		**Median**		**Runtime**	
	Trans.	**Rot.**	**Trans.**	**Rot.**	**Mean**	**Median**
FNMI ($\sigma = 2$)	7.52	0.39	0	0	143.5	88.2
RMI	94.61	1.24	16	0.5	22.6	23.9
Mean Neighbourhood MI	89.13	2.39	74	2.0	11.0	11.2
Gradient-Image MI	49.94	1.56	25	1.0	17.4	17.4
Gradient MI	80.22	2.16	65	2.0	19.5	19.4
Second-Order MI (mean neighbourhood)	105.48	1.94	25	1.5	6.1	6.2
Second-Order MI (median neighbourhood)	123.68	1.74	55	1.5	7.3	7.3
Second-Order MI (gradient value)	119.54	1.74	40	1.0	14.2	14.0
Second-Order MI (left neighbour pixel)	160.70	2.01	151	1.5	5.4	5.4
Second-Order MI (right neighbour pixel)	171.54	1.96	179	1.5	5.2	5.2
Standard MI (32 bins)	115.43	1.90	42	1.5	3.9	4.0

Table 2. Registration error for FNMI and RMI

Method	Mean		Median		Runtime	
	Trans.	Rot.	Trans.	Rot.	Mean	Median
FNMI ($\sigma = 2$ and $\sigma = 4$)	5.48	0.47	2	0	223.2	146.6
RMI ($r = 3$)	14.92	0.50	5	0.5	85.2	85.7
RMI ($r = 4$)	11.13	0.52	5	0.5	149.4	150.3
RMI ($r = 5$)	6.91	0.52	6	0.5	259.2	257.8

Table 3. Grading results for registration

Method	Grading Results				
	Excellent	V.Good	Good	Weak	Failed
FNMI ($\sigma = 2$)	30	92	8	1	4
Regional MI	11	41	11	6	66
Mean Neighbourhood MI	3	9	5	3	115
Gradient-Image MI	8	22	27	12	66
Gradient MI	1	14	10	8	102
Second-Order MI (mean neighbourhood)	10	32	9	14	70
Second-Order MI (median neighbourhood)	7	30	5	10	83
Second-Order MI (gradient value)	10	37	12	8	68
Second-Order MI (left neighbour pixel)	6	23	10	8	88
Second-Order MI (right neighbour pixel)	7	21	6	6	95
Standard MI (32 bins)	8	24	16	11	76
FNMI ($\sigma = 2$ and $\sigma = 4$)	38	88	9	0	0
RMI ($r = 3$)	23	85	21	0	6
RMI ($r = 4$)	29	85	19	0	2
RMI ($r = 5$)	37	86	12	0	0

to assess exactly how well this has improved registration using mean and median error results alone. Therefore we also provide qualitative assessment to determine the number of successful registrations shown in Table 3.

Figure 3 shows an example of a difficult registration. The result obtained using our original method is the failed registration, however FNMI with $\sigma = 2$ and $\sigma = 4$ can successfully register the images. Both methods give the global maximum at the correct registration, however using the additional features reduces the number of local maxima that can trap the search optimisation, meaning that the search is more likely to converge to the correct solution. What is interesting to note here is that if using traditional Mutual Information, the failed registration result gives MI as 0.4958 and the correct registration gives MI as 0.4672, and so the search optimisation would fail to find the correct registration, whether using a search optimisation or an exhaustive search. This shows that Mutual Information can not be relied on as a robust similarity measure.

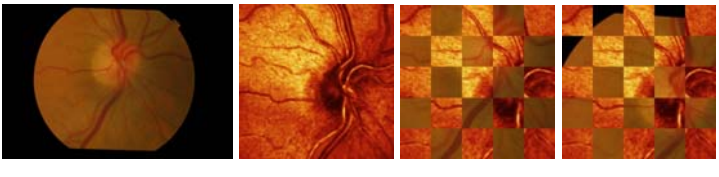


Fig. 3. Left-to-right: Fundus image, SLO image, failed and successful registration

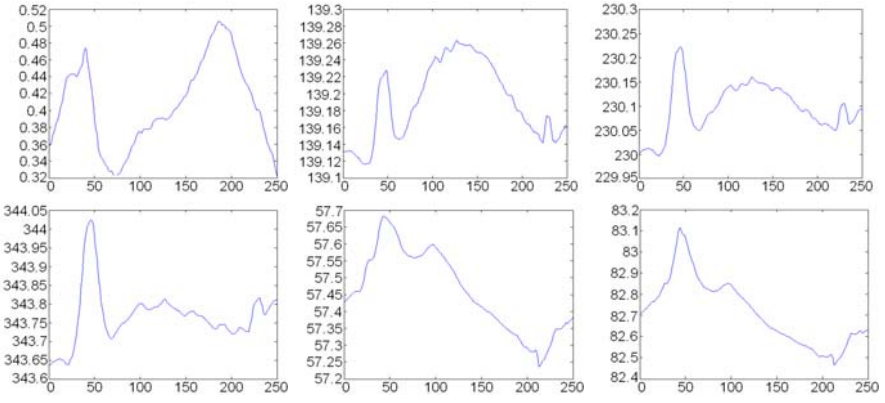


Fig. 4. Surface plots of a difficult registration. Top Row: MI (32 bins), RMI ($r = 3$), RMI ($r = 4$). Bottom Row: RMI ($r = 5$), FNMI ($\sigma = 2$) and FNMI ($\sigma = 2$ and $\sigma = 4$).

Figure 4 shows each similarity measure versus X-translation for registration of our difficult image example. In this example, both MI and RMI (when $r = 3$) fail as the true registration does not give the global maximum in our surface function. RMI ($r = 4$ and $r = 5$) both give the correct registration at the global maximum, although there exist other local maxima that could easily trap the search optimisation. FNMI peaks at the correct registration, has much fewer local maxima, and provides a smooth surface that converges well to the solution. This example considers only X-translation purely to provide clearer visualisation. If considering all possible transformations then it becomes more apparent that both the simiarity measure and the search optimisation can greater affect the performance of registration.

4 Discussion

Our study is focused on the registration of multi-modal retinal images. We propose to combine multi-scale derivatives with neighbourhood intensities, and incorporate these into a MI framework. Our method achieves very accurate results for our registration problem compared to existing registration methods.

Our method extends the work of Russakoff, who proposed RMI. In RMI, only neighbourhood intensities are included in registration. What we found with RMI is that when using the direct neighbourhood (a pixel and its 8 neighbours, for each image; giving 18 points), RMI fails to consistently give good registration results. This can be improved by increasing the radius of the neighbourhood, however the number of points required for each vector quickly becomes very large. In difficult registration cases, we found RMI can register these to a good standard when using a neighbourhood radius of 5 pixels (requiring 242 points in each column vector). The same image can be registered using our method using two feature images (requiring 54 points in each vector). This significant

reduction of data points becomes even more apparent in the covarience matrix, where RMI would have a matrix C of size 242×242 (58564 elements) compared to a matrix of size 54×54 (2916 elements) in our method. This leads to much less computation time required for this step of the algorithm. Our method does require computing feature derivatives however which takes additional time. We experimented with two approaches: computing the features first and then transforming the complete image stack, or performing transformations on the single intensity images and computing features only when performing our registration comparison. The second approach gave slight improvement to runtime and so was adopted, although both of these approaches could be performed much faster by exploiting the capabilities of the GPU, giving a robust similarity measure that would also be efficient for fast image registration.

References

1. Viola, P.A., Wells, W.M.: Alignment by maximization of mutual information. In: ICCV, pp. 16–23 (1995)
2. Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G.: Automated multimodality image registration based on information theory. In: Proc. Int. Conf. Information Processing in Medical Imaging, pp. 263–274 (1995)
3. Mellor, M., Brady, M.: Phase mutual information as a similarity measure for registration. Medical Image Analysis 9(4), 330–343 (2005)
4. Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A.: Image registration by maximization of combined mutual information and gradient information. IEEE Trans. Med. Imaging 19(8), 809–814 (2000)
5. Dowson, N.D.H., Bowden, R., Kadir, T.: Image template matching using mutual information and NP-windows. In: ICPR, vol. (2), pp. 1186–1191 (2006)
6. Rueckert, D., Clarkson, M., Hill, D., Hawkes, D.: Non-rigid registration using higher-order mutual information. In: Medical Imaging: Image Processing, pp. 438–447 (2000)
7. Beijing, C., JunLi, L., Gang, C.: Study of medical image registration based on second-order mutual information. In: IEEE International Conference on Multimedia & Expo, pp. 956–959 (2007)
8. Kubecka, L., Jan, J.: Registration of bimodal retinal images - improving modifications. In: IEEE Int. Conf. on Engineering in Medicine and Biology, pp. 1695–1698 (2004)
9. Russakoff, D.B., Tomasi, C., Rohlfing, T., Maurer, C.R.: Image similarity using mutual information of regions. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3023, pp. 596–607. Springer, Heidelberg (2004)
10. Yang, C., Jiang, T.-Z., Wang, J., Zheng, L.: A neighborhood incorporated method in image registration. In: Yang, G.-Z., Jiang, T.-Z., Shen, D., Gu, L., Yang, J. (eds.) MIAR 2006. LNCS, vol. 4091, pp. 244–251. Springer, Heidelberg (2006)
11. ter Haar Romeny, B.M., Florack, L.M.J., Koenderink, J.J., Viergever, M.A.: Scale space: Its natural operators and differential invariants. In: Colchester, A.C.F., Hawkes, D.J. (eds.) IPMI 1991. LNCS, vol. 511, pp. 239–255. Springer, Heidelberg (1991)
12. Shannon, C.: A mathematical theory of communication. Bell System Technical Journal 27, 379–423, 623–656 (1948)
13. Lagarias, J.C., Reeds, J.A., Wright, M.H., Wright, P.: Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM Journal of Optimization 9(1), 112–147 (1998)

Multimodal Image Registration by Information Fusion at Feature Level

Yang Li and Ragini Verma

Department of Radiology, University of Pennsylvania, Philadelphia, PA, USA, 19104
{Yang.Li,Ragini.Verma}@uphs.upenn.edu

Abstract. This paper proposes a novel multimodal image registration method which can fully utilize the multimodal information and result in a more accurate unified deformation field. Different from the existing methods which fuse the information at the image/intensity level, the proposed method fuses the multimodal information at the feature level through Gabor wavelets transformation. At this level, complementary and redundant information is distinguished reliably and efficiently, and then combined and removed respectively. Experiments on both simulated and real T1+DTI image sets illustrate that the proposed method can effectively incorporate better characterization for white matter (WM) from the DTI and for gray matter (GM) from the T1 image and lead to a more accurate and efficient multimodal image registration which paves the way for the subsequent multimodal population-based studies.

1 Introduction

Clinical studies routinely acquire different magnetic resonance imaging (MRI) protocols such as structural images (T1, T2, FLAIR, mPRAGE) and diffusion imaging such as Diffusion Tensor Imaging (DTI). Each of these modalities provides some unique characterization of tissue. For instance, structural images generate contrast between the gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF). DTI captures the directional microstructural information within WM as a tensor, which complements the missing orientation information of structural images. Multimodal image registration has become an active research topic, because 1) it combines complementary information from different modalities and 2) the deformation field obtained is common to all the modalities. When the spatially normalized images are used for statistics, using the same deformation field makes them more comparable than using separate registrations for each of the modality which usually generates inconsistent deformation fields. It is also efficient, as registrations on all modalities are produced at the same time. In [1], Park *et al* proposed a deformable multimodal image registration method using multichannel demons [2], in which T2 and DTI were registered as a vector image. In this straightforward method, although every image channel is incorporated, they are assigned equal importance in the image matching. Or in other words, there is no separation between complementary and redundant information. More recently, in [3], multimodal image consisting of T1, DTI and

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 624–631, 2009.

cerebral blood flow (CBF) were registered by defining the multimodal similarity metric as a weighted summation of each modality's similarity metric. Although spatially constant weights were used in the experiments, the author also suggested that a weighting function $\omega_i(\mathbf{x})$ defined on the image domain should generate better results. In Studholme's work [4], DTI was incorporated into the mapping between T1 as an constraint. Multichannel mutual information (MI) was used to match the multimodal image. Due to the impractical number of histogram bins required by the multichannel MI of all 7 image channels (T1 and 6 tensor components), a simplification of only using two tensor components, D_{xx} and D_{yy}, was adopted based on the fact that the different diffusion directions is relatively un-correlated. For other multimodal image with large number of modalities, multichannel MI remains computationally expensive.

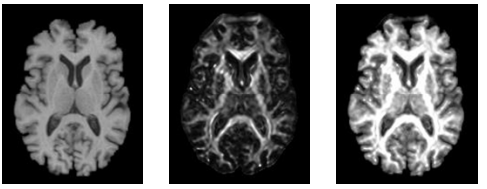
The challenging problem in multimodal image registration is to fuse the multimodal information and find a unified deformation field that can be applied to spatially normalize every image channel. The meaning of 'fuse' is two-fold here, merging the complementary (to be more accurate) and removing the redundant (to be more efficient). The main drawback of dealing with the information fusion problem at image/intensity level is that, it is difficult to extract and distinguish between complementary and redundant information, because the stronger signal (higher intensity value) does not necessarily indicates a stronger characterization ability of one modality. This ability is proportional to the contrast the modality can provide. Inspired by the ideas of multiscale decomposition (MSD)-based information fusion scheme [5], which have been successfully used in image fusion [6], in this paper, we propose a novel multimodal image registration algorithm which fuses the multimodal information at feature level. By designing a matching metric through Gabor wavelets transformation, the proposed registration method can incorporate and enhance complementary information while eliminating the redundancies between the different channels. Therefore, this leads to a more accurate registration at a much lower computational cost. In experiments, we apply the new algorithm to T1+DTI image and obtain more accurate registration than using either of the two modalities individually.

2 Methods

Before describing the proposed registration algorithm, we will illustrate the basic idea of MSD-based information fusion scheme through an example.

2.1 MSD-Based Information Fusion

An image $\mathbf{I}$ can be decomposed by discrete wavelets transform (DWT) into a series of wavelets coefficients $c^{\mathbf{I}}_{k,l,m,n}$, where k, l and (m,n) indicate the decomposition level, the frequency band and the position in a frequency band, respectively. Because $\|c^{\mathbf{I}}_{k,l,m,n}\|$ reflects the local energy of the output of the filter banks and is a good indicator of the characterization ability of each modality at the specific scale, frequency band and position, the simplest but efficient information fusion rule is the *choose-max* scheme [5]. That is, to fuse two image



(a) (b) (c)

Fig. 1. An example of image fusion by the choose-max scheme. (a) T1 image; (b) FA image; (c) Fused image.

X and **Y**, we can get the wavelets-based MSD representation of the fused image **Z** by only keeping the coefficient with the largest absolute value. After that, an inverse DWT can be performed to finally get the fused image **Z**. As an example, we use a pair of co-registered T1 and fractional anisotropy (FA) [7] of the same subject to show the performance of the above fusion scheme. As shown in Fig. 1, the fused image gives good characterization on both WM and cortex which are not available in the input T1 and FA, individually.

2.2 Choose-Max Metric for Matching

Using the similar MSD-based Choose-Max information fusion scheme, we firstly define a dissimilarity metric for the multimodal image registration. However, since the objective of information fusion in registration is to fully utilize the multimodal information for a better matching criteria, but not to get a fused image explicitly, we use Gabor wavelets [8] for the multiscale analysis, instead of DWT. Gabor wavelet transformation has been shown to be optimal in the sense of minimizing the joint uncertainty in space and frequency, and has been widely used for feature extraction [9]. Although the registration is for 3D image, to alleviate the computational cost, we use 3 perpendicular (axial, coronal and sagittal) 2D Gabor filter banks to extract the features. A 2D Gabor filter can be viewed as a sinusoidal plane of particular frequency and orientation, modulated by a Gaussian envelope: $G(x, y) = s(x, y)g(x, y)$, where $s(x, y)$ is a complex sinusoid: $s(x, y) = \exp[-j2\pi(u_0x + v_0y)]$ and $g(x, y)$ is a 2D Gaussian envelope:

$$g(x, y) = \frac{1}{\sqrt{2\pi}\sigma_x\sigma_y} \exp[-\frac{1}{2}(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2})] \tag{1}$$

σ_x and σ_y characterize the spatial extent and bandwidth of g along the respective axes, u_0 and v_0 are the shifting frequency parameters in the frequency domain. Using $G(x, y)$ as the mother wavelet, the Gabor wavelets, a class of self-similar functions can be obtained by appropriate dilations and rotations of $G(x, y)$ through: $G_{m,n}(x, y) = a^{-m}G(x', y')$, where $x' = a^{-m}(x\cos\theta + y\sin\theta)$, $y' = a^{-m}(x\sin\theta + y\cos\theta)$, $a > 1$, $\theta = n\pi/O$, $m = 1 \cdots S$ $n = 1 \cdots O$. O indicates the number of orientations and S the number of scales in the multiresolution decomposition. These parameters can be set according to [9] to reduce the redundant information (caused by the nonorthogonality of the Gabor wavelets) in

the filtered images. Given an image $I(x, y)$, the Gabor transform with orientation n and scale m can be computed as:

$$F_{m,n}(x, y) = \int I(x_1, y_1) G^*_{m,n}(x - x_1, y - y_1) dx_1 dy_1 \tag{2}$$

where $*$ indicates the complex conjugate.

Suppose we have two multimodal image $\mathbf{I} = [I_1, \cdots, I_M]$ and $\mathbf{J} = [J_1, \cdots, J_M]$, in which $I_1, \cdots, I_M$ and $J_1, \cdots, J_M$ are two collections of co-registered images generated by M different modalities. The Gabor wavelet transformation of the ith modality image I_i and J_i with orientation n and scale m are $F^{I_i}_{mn}$ and $F^{J_i}_{mn}$, respectively. Using the choose-max scheme, the dissimilarity between two voxel $\mathbf{I}(\mathbf{x_1})$ and $\mathbf{J}(\mathbf{x_2})$ can be measured as:

$$D_1(\mathbf{I}(\mathbf{x_1}), \mathbf{J}(\mathbf{x_2})) = \sqrt{\sum_{m=1}^{S} \sum_{n=1}^{O} \sum_{i=1}^{M} w_{m,n,i} (||F^{I_i}_{m,n}(\mathbf{x_1})|| - ||F^{J_i}_{m,n}(\mathbf{x_2})||)^2} \tag{3}$$

where

$$w_{m,n,i} = \begin{cases} 1 & \text{if } i = \underset{j=1\cdots M}{\operatorname{argmax}} ||F^{I_j}_{m,n}(\mathbf{x_1})|| \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

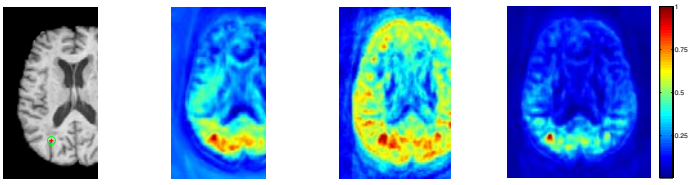
Using D_1 on two T1+FA image sets, the similarity between a reference point on the template and other points on a subject can be computed and shown in Fig. 2(b) as a similarity map.

2.3 Max-Index Metric for Matching

To utilize the information which is not included in (3) without too much computational cost, we can design another metric using the max indices which we already have. Let $p^{\mathbf{I}}_{m,n}(\mathbf{x})$ denote the label of the modality which generates the strongest Gabor feature at scale m and orientation n on $\mathbf{I}(\mathbf{x})$. Arrange these labels into a vector form $\mathbf{p^I}(\mathbf{x}) = [p^{\mathbf{I}}_{1,1}(\mathbf{x}), \cdots, p^{\mathbf{I}}_{S,O}(\mathbf{x})]$. Then $\mathbf{p^I}(\mathbf{x})$ is a very informative fingerprint of voxel $\mathbf{I}(\mathbf{x})$. Let $L(\mathbf{I}(\mathbf{x_1}), \mathbf{J}(\mathbf{x_2}))$ be the number of equal labels between $\mathbf{p^I}(\mathbf{x_1})$ and $\mathbf{p^J}(\mathbf{x_2})$. Thus, L is a similarity metric with an increasing resolution when S and O increase. Here, the 3 perpendicular 2D Gabor filter banks each has 4 scales and 6 orientations. Therefore, L has the ability to differentiate the brain tissue into $4 \times 6 \times 3 = 72$ classes. Fig. 2(c) shows the similarity map generated by metric L. As we can see, at this resolution level, only L can give considerable discriminatory power. More importantly, this is also a clear evidence showing the correctness and effectiveness of the choose-max fusion rule. Because if it is not reliable, we can not get such a specific similarity map by using only L. By combining L with (3), the metric for multimodal image matching is finally defined as:

$$D(\mathbf{I}(\mathbf{x_1}), \mathbf{J}(\mathbf{x_2})) = [1 - \frac{L(\mathbf{I}(\mathbf{x_1}), \mathbf{J}(\mathbf{x_2}))}{3SO + 1}] \cdot D_1(\mathbf{I}(\mathbf{x_1}), \mathbf{J}(\mathbf{x_2})) \tag{5}$$

The discriminative ability of metric D is shown in Fig. 2(d).



(a) (b) (c) (d)

Fig. 2. Similarity map generated by different metrics. (a) shows the reference point (denoted by a red cross) on T1. (b) is the similarity map generated by D_1. (c) is the similarity map generated by metric L (normalized to [0 1]). (d) is the similarity map generated by metric D. Red indicates high similarity.

2.4 Deformable Registration of Multimodal Image

With the dissimilarity metric defined in (5), the problem of registering multimodal image $\mathbf{I}(\mathbf{x})$ to multimodal image $\mathbf{J}(\mathbf{x})$, can be defined as finding a optimal transformation $h(\mathbf{x})$ which minimizes

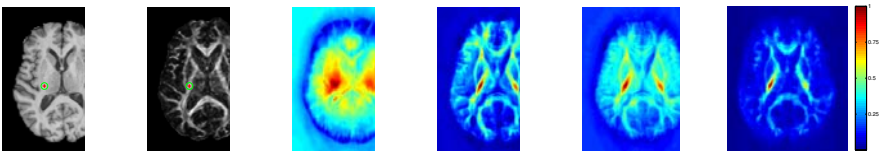
$$\int_{\mathbf{x_1}\in\Phi_\mathbf{I}} D(\mathbf{I}(\mathbf{x_1}),\mathbf{J}(h(\mathbf{x_1})))d\mathbf{x_1} + \int_{\mathbf{x_2}\in\Phi_\mathbf{J}} D(\mathbf{I}(\mathbf{x_1}),\mathbf{J}(h^{-1}(\mathbf{x_2})))d\mathbf{x_2} + \alpha \int_{\mathbf{x_1}\in\Phi_\mathbf{I}} \|\nabla^2 h(\mathbf{x})\| d\mathbf{x_1} \tag{6}$$

By hierarchically optimizing (6) through the deformation model on the lines of the HAMMER algorithm [10], we develop a multimodal image registration algorithm. To select the active points (which drive the registration), we also adopt the choose-max criterion. After the edge map [11] of each channel is computed, a fused edge map is obtained using the choose-max scheme. On this fused map, the portion of points with higher edge intensity are selected as the active points. After that, in a hierarchical manner, by adding more and more active points to guide the deformation, this algorithm generates the final spatial transformation by concatenating the hierarchical sequence of piecewise smooth transformations obtained at each stage. Since DTI is one of the modalities in our experiments, the obtained deformation field is also used to determine the tensor reorientation by using a spatially adaptive procedure that estimates the underlying fiber orientation [12].

3 Experiments and Results

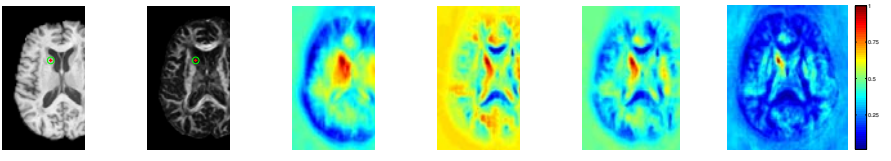
3.1 Discriminatory Ability of the Metric

The discriminatory ability of the proposed metric is firstly tested by comparing the similarity maps generated by different metrics. To make the comparison clear, we use multimodal image only consists of T1 and FA. In preprocessing, T1 is affinely registered to B_0 of DTI and re-sampled to have the same size ($256 \times 256 \times 70$) and resolution ($0.9375 \times 0.9375 \times 2.5mm$). The Gabor filter bank is set to have 6 orientations and 4 scales. Three other metrics are used



(a) (b) (c) (d) (e) (f)

Fig. 3. Comparison of the similarity maps of a reference point on WM. (a) and (b) show the reference point (denoted by a red cross) on T1 and FA, respectively; (c) and (d) are similarity map only using T1 and FA, respectively; (e) is the similarity map using both T1 and FA without fusing; (f) is the similarity map given by the new metric defined in (5).



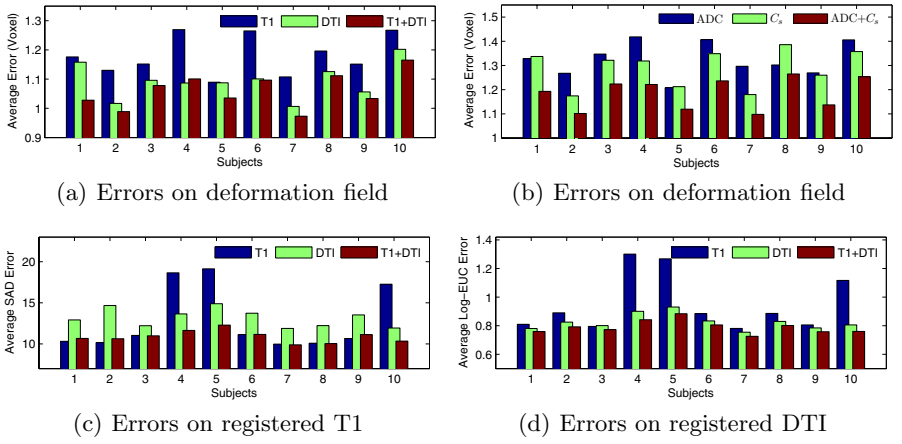
(a) (b) (c) (d) (e) (f)

Fig. 4. Comparison of the similarity maps of a reference point on GM. Legends are the same as those in Fig. 3.

for comparison: metric only using the Gabor features from T1 or FA and the metric using features from both T1 and FA but without fusing (as in (3) with $w_{m,n,i} = 1, \forall m, n, i$). As expected, for a reference point on WM, the metric only using FA features (Fig. 3(d)) is more discriminative than the metric only using T1 features (Fig. 3(c)). And similarly, the metric only using T1 features (Fig. 4(c)) is more discriminative on GM than the metric only using FA features (Fig. 4(d)). Although for reference point on GM, the metric using both T1 and FA without fusing improves the discriminatory power (Fig. 4(e)), by comparing Fig. 3(e) with Fig. 3(d), we also found that this metric actually degenerates the discriminatory ability on WM. In comparison, the proposed new metric (defined in (5)) correctly utilizes the information from both channels and gives the most discriminative result on both WM (Fig. 3(f)) and GM (Fig. 4(f)).

3.2 Registration of Simulated and Real Images

Ten simulated DTI+T1 image set are generated by applying ten simulated deformation fields on a template DTI+T1 image set (with the same size, resolution and preprocessing as used in Sect. 3.1). These ten simulated DTI+T1 image are then registered back to the template using the Gabor features extracted from T1 and 5 DTI-derived scalar images (FA, apparent diffusion coefficient (ADC) [7], C_s (sphericity), C_p (oblateness) and C_l (prolateness) [13]). For comparison, 3 registrations were computed: the proposed method and the other two methods which only use features from one modality, either T1 or DTI. The deformation fields generated by the 3 methods are then compared with the simulated



(a) Errors on deformation field (b) Errors on deformation field

(c) Errors on registered T1 (d) Errors on registered DTI

Fig. 5. Comparison of the registration accuracy

deformation fields (ground truth). The results are summarized in Fig. 5(a), in which the proposed method, by utilizing the information from both modality, outperforms the other two methods on most of the subjects. The performances of the three methods are also compared on ten real T1+DTI images. After warping both T1 and DTI using the 3 deformation fields generated by the 3 methods, the registration accuracies were compared on both the registered T1 and the registered DTI based on the average Sum Absolute Difference (SAD) and Log-Euclidean (Log-EUC) [14] metric, respectively. In both of these comparisons (as summarized in Fig. 5(c) and Fig. 5(d)), the proposed method generates the smallest error on most of the subjects, especially the 3 difficult subjects (4, 5 and 10), on which the single modality methods generate relatively high registration error. This suggests that fusing the information from multi-modalities can help to robustly obtain accurate registration. We also found that, on some subjects (simulated subject 4 and real subject 1, 2 and 9), the proposed method does not give the best results (but difference is slight). One of the possible reasons that may cause registration error for the proposed method is the registration error between T1 and DTI in the preprocessing step. Therefore, to exclude the influence of this factor, another experiment was also conducted on the simulated subjects only using ADC and C_s (they are inherently co-registered). As shown in Fig. 5(b), the proposed method consistently gives the best result in this experiment. And because ADC and C_s have less redundant and more complementary information, the proposed method gains more improvement. In the future, more advanced inter-modality registration algorithm will be adopted to improve the intra-subject registration between T1 and DTI.

4 Summary

In this paper, we proposed a multimodal image registration framework based a new metric which combines complementary and removes redundant information

from different modalities at the feature level. Experiments on both simulated and real multimodal images (DTI+T1) illustrate that the proposed method can effectively fuse the information from different modalities and result in a more accurate unified registration. In the future, we plan to explore more advanced fusion schemes of the Gabor features. As an application, we also plan to apply this method on clinical studies for joint comparative statistics on T1 and DTI.

Acknowledgments. The authors and the work was supported by the National Institute of Health via grant R01-MH-079938. We also acknowledge the help from the work of Dr. Dinggang Shen and Yangming Ou.

References

1. Park, H.J., Kubicki, M., Shenton, M.: Spatial normalization of diffusion tensor MRI using multiple channels. Neuroimage 20, 1995–2009 (2003)
2. Thirion, J.P.: Non-rigid matching using demons. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 245–251 (1996)
3. Avants, B.B., Duda, J.T., Zhang, H., Gee, J.C.: Multivariate normalization with symmetric diffeomorphisms for multivariate studies. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 359–366. Springer, Heidelberg (2007)
4. Studholme, C.: Incorporating DTI data as a constraint in deformation tensor morphometry between T1 MR images. In: Information Processing in Medical Imaging (IPMI), pp. 223–232 (2008)
5. Zhang, Z., Blum, R.S.: A categorization of multiscale-decomposition-based image fusionschemes with a performance study for a digital camera application. Proceedings of the IEEE 87(8), 1315–1326 (1999)
6. Alfano, B., Ciampi, M., De Pietro, G.: A wavelet-based algorithm for multimodal medical image fusion. In: Semantic Multimedia, pp. 117–120 (2007)
7. Bassera, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative-diffusion-tensor MRI. Journal of Magnetic Resonance 111(3), 209–219 (1996)
8. Daugman, J.: Complete discrete 2d gabor transforms by neural networks for image analysis and compression. IEEE Transactions on Acoustics 36(7), 1169–1179 (1988)
9. Ma, W., Manjunathi, B.: Texture features for browsing and retrieval of image data. IEEE Trans. Pattern Anal. Machine Intell. 18(8), 837–842 (1996)
10. Shen, D.: Image registration by local histogram matching. Pattern Recognition 40, 1161–1171 (2007)
11. Canny, J.: A computational approach to edge detection. IEEE Trans. Pattern Anal. Machine Intell. 8(6), 679–698 (1986)
12. Xu, D., Mori, S., Shen, D., Zijl, P.C.M.V., Davatzikos, C.: Spatial normalization of diffusion tensor fields. Magnetic Resonance in Medicine 50(1), 75–182 (2003)
13. Westin, C.F., Maier, S.E., Mamata, H., Nabavi, A., Jolesz, F.A., Kikinis, R.: Processing and visualization of diffusion tensor MRI. Medical Image Analysis 6(2), 93–108 (2002)
14. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine 56, 411–421 (2006)

Accelerating Feature Based Registration Using the Johnson-Lindenstrauss Lemma*

Ayelet Akselrod-Ballin[1], Davi Bock[2], R. Clay Reid[2], and Simon K. Warfield[1]

[1] Computational Radiology Laboratory, Children's Hospital, Harvard Medical School, Boston, USA
[2] Department of Neurobiology, Harvard Medical School, USA

Abstract. We introduce an efficient search strategy to substantially accelerate feature based registration. Previous feature based registration algorithms often use truncated search strategies in order to achieve small computation times. Our new accelerated search strategy is based on the realization that the search for corresponding features can be dramatically accelerated by utilizing Johnson-Lindenstrauss dimension reduction. Order of magnitude calculations for the search strategy we propose here indicate that the algorithm proposed is more than a million times faster than previously utilized naive search strategies, and this advantage in speed is directly translated into an advantage in accuracy as the fast speed enables more comparisons to be made in the same amount of time. We describe the accelerated scheme together with a full complexity analysis. The registration algorithm was applied to large transmission electron microscopy (TEM) images of neural ultrastructure. Our experiments demonstrate that our algorithm enables alignment of TEM images with increased accuracy and efficiency compared to previous algorithms.

1 Introduction

Image registration is a fundamental process in medical imaging applications aimed at establishing spatial correspondences between images [1]. Registration algorithms must satisfy demanding requirements of speed, robustness and accuracy depending on the specific application involved. Alignment of Electron Microscopy (EM) images of neural tissue involves consideration of the high resolution and overwhelming size of the data, the large amount of detail, the acquisition artifacts and the deformation induced by the intrinsic deformation of the slices [2]. Thus, the reconstruction of neural circuitry from EM remains a substantial and challenging problem [3].

Existing algorithms for registration can be broadly classified into approaches that directly operate on image intensities or to feature based alignments seeking to identify features that should be aligned and an optimal transformation

* This investigation was supported in part by a research grant from CIMIT, grants RG 3478A2/2 and RG 4032A1/1 from the NMSS, and by NIH grants R03 CA126466, R01 RR021885, R01 GM074068, R01 EB008015 and P30 HD018655.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 632–639, 2009.

that brings them into alignment [1]. Similar to previous registration methods for alignment of microscopy images our method relies on matching local image patches across successive slices. A recent study [2] addressed matching as part of a complete algorithm for assembling 3D volumes from EM data. This approach first identifies feature descriptors based on a gradient vector pyramid and then exploits these features to match adjacent slices. An approach for 3D reconstruction based on a block matching strategy was presented by [4] where the local displacements were utilized to robustly estimate a global transformation. [5] developed a solution for 3D reconstruction of a series of EM images based on the finite support properties of the cubic B-splines, where the initial estimate for the affine registration was based on the technique described in [4].

In this work we presents a novel algorithm designed for alignment of large cross-sectional slices of EM by introducing a search strategy which to our knowledge has not been used for image registration before. The key advance we propose is based on the realization that the search for corresponding features can be dramatically accelerated by utilizing Johnson-Lindenstrauss (JL) [6] dimension reduction. This enables us to consider more regions in the images and to consider all of the potential correspondences, ensuring that the search does not overlook the correct set of correspondences, and thus dramatically improves the robustness and accuracy of the registration. Combining this together with recent advances in approximate nearest neighbor (ANN) techniques [7] for the search strategy reduces the computational complexity significantly.

The paper is organized as follows. Section 2 describes the algorithm, section 3 presents experimental results and conclusions are presented in section 4.

2 Methods

The volume reconstruction is obtained by composing pairwise 2D alignments of consecutive slices by taking as reference the middle of the stack. Thus we focus on the 2D registration of successive sections, although the algorithm can be used for 3D registration as well. The input includes the fixed scene image I_S and a moving model image I_M. Our aim is to find the transformation T aligning the scene with the model. The algorithm consists of three main steps. **In the first step**, we extract image patches, also called blocks, fragments or templates [4], as the features for matching. The patches are rectangular sub-images of $d = 10^4$ voxels (i.e. 100×100 sized image regions) extracted from the scene and model images. Then using JL-embedding [6] with the random projections approach by Achlioptas [8] we generate a low dimensional representation of these patches. **In the second step** the patches of reduced dimension are compared across successive slices. So that, given a projected scene and model patch we search for correspondences based on the Euclidean distance between the patches. **Finally**, we construct two sets of points based on the center coordinates of the projected scene and model patches. We compute the transformation with an extension to the Expectation Maximization Iterative Closest Point (EM-ICP) [9] algorithm which includes the similarity measure of the patches. Table 1 presents an outline of the algorithm proposed.

Table 1. Outline of the algorithm

Given a pair of successive slices, a fixed (scene) and moving (model) image:
Step 1: JL dimension reduction of features: State of the art feature matching approaches to registration commonly perform feature detection by correlation, followed by a truncated search and a transform estimation. In contrast with previous methods we first extract small patches at different locations for both images and then the features are projected on to a lower dimension based on the JL-Lemma.
Step 2: Search for correspondences: The projected features extracted from both images are compared based on the correlation measure or Euclidean distance.

-2a. Brute Force search strategy.
-2b. Approximate nearest neighbor (ANN) search where each near neighbor is reported with a certain probability. For this step there are several solutions such as tree based search or locality sensitive hashing [10].

Step 3: Transformation estimation: Apply extended EM-ICP [9] approach based on the corresponding features between the images.

2.1 Step 1: Johnson-Lindenstrauss (JL) Dimension Reduction

The essential property enabling accelerated search has been the realization that efficient high dimensional search can be achieved by creating randomized projections into low-dimensional spaces, and then using efficient low-dimensional search. The JL Lemma [6] asserts that any set of n points in d-dimensional Euclidean space can be projected down to k-dimensional Euclidean space, where $k = O(\epsilon^{-2} \log n)$, while maintaining pairwise distances with a low distortion. Recent research [8] has shown that random projection matrices can be used for JL projection. Following this proof, given the initial set of n points in R^d, represented as an $n \times d$ matrix, where each feature-patch is represented by a row, let R be a $d \times k$ random matrix with $R(i,j) = r_{ij}$; where the independent random variables r_{ij} are : $\{1$ with probability 0.5, and -1 with probability 0.5 $\}$. Naively, the random projection can be performed by constructing a $k \times d$ random matrix; so that mapping each point takes $O(dk)$, however recent theoretic work suggests that a projection from dimension d to dimension k can be computed with $O(d)$ operations [11].

2.2 Step 2: Search for Correspondences

The transformation computation is based on maximizing the similarity of corresponding patches. Currently we utilize the brute-force approach which involves computing the distances between the patch and all the patches in the neighboring image. Our experiments compare the results of using the full-size patches and their projections. We use normalized correlation (NC) as a similarity measure since NC is invariant to linear intensity transformation and it is assumed that for small corresponding image patches across the two successive slices, the intensities are locally related by some linear intensity transformation [4]. Another

attractive property of NC is that it is equivalent to a squared Euclidean distance [12] which meets the requirements of the JL Lemma.

Recently it has been discovered that allowing approximate rather than exact search enables dramatically accelerated search [7]. In the approximate nearest neighbor (ANN) formulation, given a query point q and a specified constant $c > 0$, the algorithm returns a point whose distance from q is at most a $c = (1+\epsilon)$-factor larger from the distance of the nearest point p in P from q. This approach results in efficient algorithms which are based on data structures such as tree based search (spill trees) and locality sensitive hashing (LSH) [10]. Thus instead of performing the naive search described above, additional significant acceleration can be obtained by casting the registration problem as an ANN search problem.

2.3 Step 3: Transform Estimation

Given the set of patches represented by their center coordinates this step determines the transformation that matches a set of model and scene points. The approach exploits the Expectation Maximization (EM) scheme to optimize simultaneously for correspondences and the registration transformation. It extends the EM-ICP approach to go from point matching to patch matching. Let s_i be the points of the scene set $S \in R^2$ and m_i the points of the model patch set $M \in R^2$, with n_s and n_m determining the number of points respectively. T represents the rigid transformation from the scene to the model. The probability of a point s_i to correspond to the model points m_i is modeled by a Gaussian probability distribution. In the case of homogeneous isotropic Gaussian noise the probability is modeled by

$$p(s_i|m_j, T) = \exp(-||T * s_i - m_j||^2/2\sigma^2) \tag{1}$$

where σ represents the noise in the measurement.

The idea is to maximize $\log p(S, A|M, T)$ the log-likelihood of the data distribution where the unknown correspondences $A \in R^{n_S \times n_M}$ are considered as a hidden random variables . The algorithm starts by initialization of the transformation (T), and repeats until convergence of the two EM steps. In the E-step, T is fixed and the probability of matches $\overline{(A_T)_{ij}}$ are computed as follows (for more details see [9])

$$\overline{(A_T)_{ij}} = \frac{\overline{\pi_{ij}} \exp(-||T * s_i - m_j||^2/2\sigma^2)}{\sum_k \overline{\pi_{ik}} \exp(-||T * s_i - m_k||^2/2\sigma^2)} \tag{2}$$

In the classic EM-ICP, the prior probability of the matches is based on the uniform law: $\pi_{ij} = \frac{1}{n_M}$. In contrast, the EM-ICP-NC used here extends the prior to account for the NC similarity of the patches taking advantage of the similarity measure in addition to the spatial information. Hence, the prior is based on the normalized NC:

$$\overline{\pi_{ij}} = \frac{NC(p_i, p_j)}{\sum\limits_k NC(p_i, p_k)} \tag{3}$$

The NC-based prior $\overline{\pi_{ij}}$ is also used in the initialization step to determine the initial transformation T. In the M-step, A is fixed and the likelihood is optimized w.r.t to T.

$$T_{n+1} = \arg\max_{T}(E_A[\log P(S, A|M, T)]) \tag{4}$$

Thus, by ignoring constant terms the criterion optimized by the EM-ICP-NC yields

$$C_{A_T}(T) = \sum_{i}^{n_s}\sum_{j}^{n_m}\overline{(A_T)_{ij}}\log p(s_i|m_j, T) \tag{5}$$

2.4 Computational Complexity

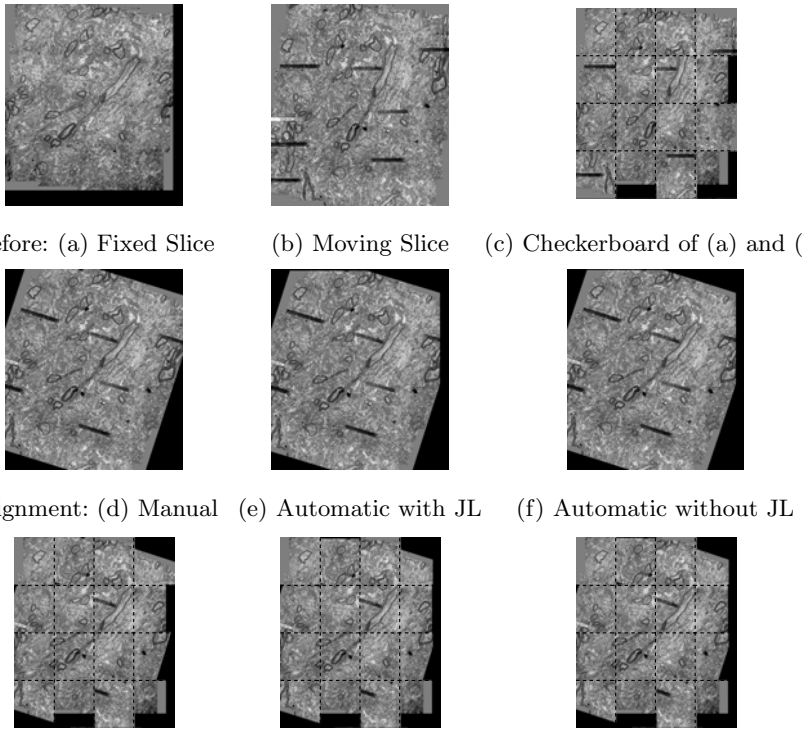
The computational complexity of the entire process is determined by the number of images ($T \approx 100$) in the data base, the image size n (we assume that the number of patterns per slice is proportional to n) and the feature dimension d. Our preliminary results were performed on slice sections of $n = 10^8 = 10^4 \times 10^4$, which were downsampled using Gaussian smoothing plus bi-linear interpolation to a size of $n = 10^6 = 10^3 \times 10^3$. Improved TEM techniques utilizing multiple camera arrays have lead to datasets of $n = 10^{10} = 10^5 \times 10^5$. The size of a typical local image patch is $d = 10^4 = (100 \times 100)$, by employing JL lemma the initial $n \times d$ matrix is projected to an $n \times k$ feature matrix (where $k = 100$ in our experiments). We focus on the search for correspondences since it is the most expensive operation of the algorithm. The naive searching method involves comparing each projected patch to all the projected patches in the neighboring image. The complexity of the naive search time per query is $O(dn)$. Thus, the overall complexity for using the full d-dimensional patches is $O(Tn^2d)$. Hence, projection to lower dimension alone reduces the complexity to $O(Tn^2 \log n)$ saving $O(d-k)$ operations (10^4 operations per slice in our experiments).

Additional significant acceleration can be obtained by employing recent advances in approximation algorithms for performing ANN. The authors in [7] have shown that the time required for retrieving similar features is reduced to $O(dn^{1/c^2})$ where for $c = 2$ this becomes $O(dn^{1/4})$. Consequently, the proportion between the naive and accelerated approaches is $O(n^{1-1/c^2})$ and for $c = 2$ this becomes $O(n^{3/4})$ which in the typical conditions of the data we are dealing with becomes $O(10^{7.5})$. Thus we claim that the proposed approach to registration is more than a million times faster than using naive approaches. We also note that the query time quoted assumes the need to project from high dimension to low dimension before doing the query. However, since our patches participate as both search points and query points, we can benefit from pre-computing the projection of all the points ahead of the query, which reduces the query cost by $O(d)$. Recall d is $O(10^4)$ in our case, so this is a substantial improvement.

3 Experiments and Results

The algorithm was tested on a series of EM images of the lateral geniculate nucleus of a ferret. Each image is about 10000×10000 pixels large with a pixel

resolution of $3nm$ and a slice thickness of $60nm$. We evaluated our experiments on 114 slices which were manually preprocessed to correct or exclude images with significant artifacts or severe rotations. The approach was validated by comparison to a manual ground truth registration. The manual registration denoted as (T^*), was performed by manually selecting corresponding points in a pair of consecutive images and computing the pairwise transformation based on Horn's method [13]. Given a set of corresponding points in two systems, Horn's method finds the closed form solution to the "least square" problem relating these points. We performed two sets of experiments. The first set of experiments was obtained with the JL projection of features while the second set was obtained without the projection. Both experiments compare the results obtained by three automatic approaches including our EM-ICP-NC approach (see Sec. 2.3), the classic EM-ICP [9] and Horn's [13] approach. The L_2 difference between the transformations



Before: (a) Fixed Slice (b) Moving Slice (c) Checkerboard of (a) and (b)

Alignment: (d) Manual (e) Automatic with JL (f) Automatic without JL

Checkerboard: (g) Manual(h) Automatic with JL (i) Automatic without JL

Fig. 1. Successive pair of slices before and after alignment. (a) and (b) present the fixed and moving image before alignment and (c) shows their checkerboard composite. (d),(e),(f) show the alignments results of the manual, and automatic algorithms with JL projection and without projection respectively. (g),(h),(i) demonstrate the checkerboard of the fixed image in (a) and the aligned moving image in (d), (e) and (f) respectively.

Table 2. Comparing our results to the manual transformation and to two other automatic approaches. Our algorithm has higher accuracy than previously described methods. The advantage is obtained both with and without the JL projection.

Dimension	EM-ICP-NC	EM-ICP	Horn
k **(with JL projection)**	5.41 ± 3.32	45.024 ± 28.56	93.63 ± 56.9
d **(without JL projection)**	4.93 ± 2.94	20.55 ± 26.74	66.86 ± 53.6

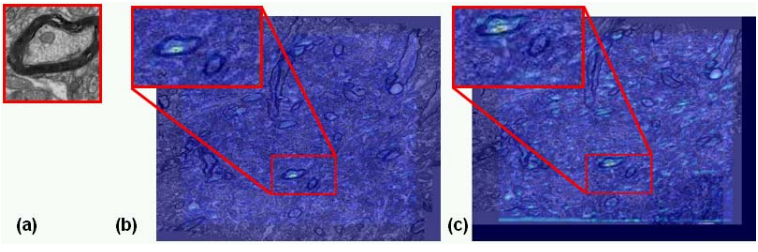


Fig. 2. Effectiveness of template matching for identifying correspondences. The fusion of the correlation map computed for the patch (a) with the original slice, from which the patch was extracted (b) and with the successive slice (c). The maxima regions in both slices (b,c) respectively, are highlighted in red rectangles and enlarged. The red maxima values in the fusion image correspond to higher correlation values, showing that the features are a sharp local maxima of the correlation function.

of all three automatic approaches in both sets of experiments are compared to the ground truth transformation computed (T^*). Figure 1 shows the results obtained with and without the JL projection compared to the manual registration. Table 2 presents the results of both experiments. Our EM-ICP-NC automatic algorithm has a higher accuracy than previously described methods. The advantage is achieved in both set of experiments with and without the JL projection, with similar results. Thus, we conclude that the projection to low dimensional space did not reduce the ability to recognize corresponding patches.

To evaluate the effectiveness of template matching for identifying correspondences in various cases of neurobiological objects, we tested patches with different features, such as myelinated white matter, dendrites, synapses, and microtubules. The NC function was computed between the features visible in the image, with patches within the same image and with patches in a successive image. Figure 2 illustrates the results, demonstrating that the local maxima in the NC function within the same image is detected in the exact same spot, while in the next slice the detection is in the area of similar anatomy.

4 Summary

We present a novel algorithm for alignment of large EM images. To our knowledge this is the first attempt to provide the registration algorithm a low

dimensional representation of the data by utilizing the JL embedding and to demonstrate the dramatic speed up of the identification of correspondence which at the same guarantees the robustness and accuracy of the alignment. In this work we have demonstrated these contributions on real EM data and also constructed a full scheme with significant computational savings. We assessed the impact of dimensionality reduction on the transformation accuracy and shown for the first time the feasibility and effectiveness of this approach, showing that the projected features are as effective for registration as the full-dimensionality features. Future work will evaluate alternative search strategies for the acceleration of accurate correspondence estimation. This will include evaluation of data structures that support ANN search strategies, such as LSH, and spill-trees. We will explore faster variants of algorithms to project into k-dimensional space.

References

1. Maintz, J.B.A., Viergever, M.A.: A survey of medical image registration. Medical image analysis 2(1), 1–16 (1998)
2. Koshevoy, P., Tasdizen, T., Whitaker, R., Jones, B., Marc, R.: Assembly of large three-dimensional volumes from serial-section TEM. In: MICCAI (2006)
3. Smith, S.J.: Circuit reconstruction tools today. Curr. Opin. Neuro. 17(5) (2007)
4. Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3D structure from serial histological sections. IVC 19, 25–31 (2000)
5. Dauguet, J., Bock, D., Reid, C.R., Warfield, S.K.: Alignment of large image series using cubic b-splines tessellation: Application to TEM data. In: MICCAI (2006)
6. Johnson, W., Lindenstrauss, J.: Extensions of lipschitz mapping into a hilbert space. Contemp. Math. 26, 189–206 (1984)
7. Indyk, P., Motwani, R.: Approximate nearest neighbors: towards removing the curse of dimensionality. In: STOC 1998, pp. 604–613 (1998)
8. Achlioptas, D.: Database-friendly random projections:johnson-lindenstrauss with binary coins. Journal of Comp. and Sys. Sci. 66, 671–687 (2003)
9. Granger, S., Pennec, X.: Multi-scale EM-ICP: A fast and robust approach for surface registration. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2353, pp. 418–432. Springer, Heidelberg (2002)
10. Andoni, A., Indyk, P.: e^2lsh: Exact euclidean locality-sensitive hashing. Implementation (2004), http://web.mit.edu/andoni/www/LSH/index.html
11. Liberty, E., Ailon, N., Singer, A.: Dense fast random projections and lean walsh transforms. LNCS, pp. 512–522 (2008)
12. Arya, K., Gupta, P., Kalra, P., Mitra, P.: Image registration using robust m estimators. Pattern Recognition Letters 28, 1957–1968 (2007)
13. Horn, B.K.P.: Closed-form solution of absolute orientation using unit quaternions. Journal of the Optical Society of America 4, 629–642 (1987)

Groupwise Registration and Atlas Construction of 4th-Order Tensor Fields Using the $\mathbb{R}^+$ Riemannian Metric*

Angelos Barmpoutis and Baba C. Vemuri**

CISE Department, University of Florida, Gainesville, FL 32611, USA
{abarmpou,vemuri}@cise.ufl.edu

Abstract. Registration of Diffusion-Weighted MR Images (DW-MRI) can be achieved by registering the corresponding 2nd-order Diffusion Tensor Images (DTI). However, it has been shown that higher-order diffusion tensors (e.g. order-4) outperform the traditional DTI in approximating complex fiber structures such as fiber crossings. In this paper we present a novel method for unbiased group-wise non-rigid registration and atlas construction of 4th-order diffusion tensor fields. To the best of our knowledge there is no other existing method to achieve this task. First we define a metric on the space of positive-valued functions based on the Riemannian metric of real positive numbers (denoted by $\mathbb{R}^+$). Then, we use this metric in a novel functional minimization method for non-rigid 4th-order tensor field registration. We define a cost function that accounts for the 4th-order tensor re-orientation during the registration process and has analytic derivatives with respect to the transformation parameters. Finally, the tensor field atlas is computed as the minimizer of the variance defined using the Riemannian metric. We quantitatively compare the proposed method with other techniques that register scalar-valued or diffusion tensor (rank-2) representations of the DWMRI.

1 Introduction

Group-wise image registration is a challenging task in medical imaging which is related to the problem of computing an atlas, i.e. the image of the average subject from a set of co-registered subjects. There are two prevalent approaches for atlas construction. The first one is based on group-wise alignment of 3D shapes [1,2], while the second one is uses alignment of 3D image intensity maps.

In this paper we focus on the second category, and therefore we review only techniques that are based on intensity map registration. Joshi et al. [3] proposed a method for group-wise image registration and simultaneous atlas construction. In this method the atlas is formed by minimizing the distance between the displacement fields that warp the images and therefore it is not biased toward a specific subject data. The estimated atlas does not belong to the set of registered subjects unlike the method presented in [4], which perform pair-wise registration of all the subjects and select the least biased target as the atlas.

* This research was in part funded by the NIH grant EB007082 to BCV.
** Corresponding author.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 640–647, 2009.

The aforementioned methods perform scalar-valued image registration. It has been shown, however, that registration of diffusion tensor-valued images (DTI) produces more accurate alignments of fibrous tissues [5]. In this approach the tensors should be re-oriented appropriately after the warping of the DTI images in order to preserve the micro-structural geometry in the subjects. One way to avoid the tensor re-orientation is to register rotation invariant quantities or other highly structured features extracted from DTI [6]. A DTI similarity measure that uses the full information in the tensors and performs their re-orientation using locally affine transformations was employed in [7]. Furthermore, two methods for diffeomorphic non-rigid DTI registration were proposed in [8] and [9] both of which use analytic derivatives of the reorientation term in the corresponding energy functions.

All the above techniques perform pair-wise DTI registration. Multi-subject registration for DTI atlas construction was proposed in [10] by extending the scalar-image framework in [3]. Another group-wise DTI registration technique which unfolds the manifold described by the Geodesic-Loxodromes metric on diffusion tensors and produces vector-valued images that are being warped in order to estimate the DTI atlas was recently proposed in [11].

Although the methods for DTI registration and atlas construction yield richer representations than the corresponding scalar-image based techniques, they fail in regions of fiber crossings and other complex tissue geometries since 2^{nd}-order tensors cannot account for multiple peaks in the diffusivity function. This problem can be resolved by using $4^{th} - order$ tensor fields and registering them using the recently proposed method in [12]. In their work, it was shown that the alignment of $4^{th} - order$ tensor fields produces more accurate results compared to those obtained by DTI registration. This technique performs tensor comparison using Hellinger's distance, which is however defined between probabilities. Since diffusion tensors are not probabilities, Hellinger's distance is not a suitable measure, unless we perform tensor normalizations which are unnatural and we avoid in this paper. Furthermore, the method in [12] performs pair-wise tensor field registration and hence cannot be directly employed for group-wise registration or statistical atlas construction.

In this paper we present a novel method for unbiased 4^{th}-order tensor field atlas construction. Our method (significantly) generalizes the unbiased diffeomorphic scalar image atlas construction framework in [3] to the case of symmetric positive definite higher-order tensors. The atlas is computed simultaneously with the non-rigid deformation fields using a functional minimization procedure. We define a novel cost function using the Riemannian metric on positive valued functions which is a generalization of the Riemannian metric on $\mathbb{R}^+$. This metric appropriately handles the positive nature of the symmetric positive-definite high-order tensors and their re-orientation is performed analytically using the Gram-Schmidt orthogonalization process of the local Jacobian matrices. The method is validated using synthetic and real DW-MRI data from isolated human hippocampi.

The key contributions of this work are: To the best of our knowledge, this is the first report in literature for higher-order tensor field atlas construction. Our method outperforms the existing methods that register derived scalar images

or 2^{nd}-order tensor fields from DWMRI, both of which fail to accurately warp datasets with complex local tissue structures such as fiber crossings. Furthermore, we employ a novel metric based on the Riemannian geometry of positive-valued spherical functions and we show that it produces more accurate results compared to the standard Euclidean metric. Finally, our cost function has analytic derivatives with respect to the unknown transformation parameters that lead to an efficient and easily scalable implementation of our framework.

2 Riemannian Metric for Positive-Valued Real Functions

Assume $a, b \in \mathbb{R}^+$, i.e. are elements of the space of positive real numbers. The Logarithmic map at location a is given by $Log_a(x) = log(x/a)$ and corresponds to the local tangent vector toward x. Its inverse function is the Exponential map, which is given by $Exp_a(t) = exp(t)a$ and projects the tangent $t \in \mathbb{R}$ back to the space $\mathbb{R}^+$. The corresponding Riemannian distance between a and $b \in \mathbb{R}^+$ is given by the length of the tangent

$$dist(a, b) = \left| \log \frac{a}{b} \right| \tag{1}$$

which satisfies scale invariance, i.e. $dist(sa, sb) = dist(a, b)\ \forall a, b, s \in \mathbb{R}^+$, additionally to the properties of distance measures.

The Riemannian metric in $\mathbb{R}^+$ can also be used to define distances between positive-valued functions $f_a(\mathbf{x})$ and $f_b(\mathbf{x})$ $\mathbf{x} \in \Omega$ as follows: $dist^2(f_a, f_b) = \int_\Omega dist^2(f_a(\mathbf{x}), f_b(\mathbf{x}))d\mathbf{x}$. In the particular case of parametric spherical functions $d(\mathbf{g}; \mathbf{D}_1)$ and $d(\mathbf{g}; \mathbf{D}_2)$, where $\mathbf{g} \in S^2$ and $\mathbf{D}_1$ and $\mathbf{D}_2$ are the corresponding parameter vectors, the distance is given by

$$dist^2(\mathbf{D}_1, \mathbf{D}_2) = \int_{S^2} \left| \log \frac{d(\mathbf{g}; \mathbf{D}_1)}{d(\mathbf{g}; \mathbf{D}_2)} \right|^2 d\mathbf{g}. \tag{2}$$

Note that the integral in Eq. 2 is over S^2, i.e. the space of unit vectors $\mathbf{g}$. This distance function is invariant with respect to 3D rotations and scale, i.e. $dist(s\mathbf{R} \circ \mathbf{D}_1, s\mathbf{R} \circ \mathbf{D}_2) = dist(\mathbf{D}_1, \mathbf{D}_2)\ \forall s \in \mathbb{R}_+$ and $\mathbf{R} \in SO_3$.

Similarly, the distance between ordered n-tuples whose elements are positive real numbers can be defined using the Riemannian metric in $\mathbb{R}^+$. In this case the distance between $A = \{a_1, a_2, \ldots, a_n\}$ and $B = \{b_1, b_2, \ldots, b_n\}$ $a_i, b_i \in \mathbb{R}^+$ is given by $dist^2(A, B) = \sum_{i=1}^{N} dist^2(a_i, b_i)$. This can also be seen as a discrete approximation of Eq. 2 by taking $a_i = d(\mathbf{g}_i; \mathbf{D}_1)$ and $b_i = d(\mathbf{g}_i; \mathbf{D}_2)$, where $\mathbf{g}_i$ is a predefined set of vectors in S^2.

In the next section we will employ the above distance measure in order to achieve simultaneous group-wise registration and atlas construction of fields of spherical functions modeled using Cartesian tensor bases of order 4.

3 Groupwise Registration of 4^{th}-Order Tensor Fields

Cartesian tensor bases of various orders have been used for approximating physical quantities computed from DW-MRI datasets. 4^{th}-order tensors

$d(\mathbf{g};\mathbf{D}) = \sum_{i,j,k,l} D^{i,j,k,l} g_i g_j g_k g_l$ have been employed to approximate the diffusivity function in generalized diffusion tensor images [13], and the kurtosis component of the diffusion in diffusion kurtosis images [14].

In the case of 4^{th}-order generalized diffusion tensors, the diffusivity is a positive-valued function and can be computed using the parametrization in [15]. This produces fields of positive-valued spherical functions whose processing can be achieved using the Riemannian metric presented in Sec. 2.

The problem of group-wise registration of N tensor-fields and simultaneous atlas estimation can be formulated as a functional minimization problem. By using Eq. 2 the energy function to be minimized is given by

$$E(\phi_n, \mathbf{D}_\mu) = \sum_{n=1}^{N} \int_\Omega \int_{S_2} \left(\log \frac{d(\mathbf{g};\mathbf{D}_n \circ \phi_n)}{d(\mathbf{g};\mathbf{D}_\mu)} \right)^2 d\mathbf{g}dx + \sum_{n=1}^{N} \int_\Omega cost(\phi_n)dx \quad (3)$$

where $\mathbf{D}_\mu$ is the 4^{th}-order tensor coefficients of the estimated atlas, ϕ_n is the estimated deformation to be applied to the n^{th} tensor field, and $cost()$ is a cost function that adds smoothing constraints to the estimated deformations.

Note that the tensor coefficients are dependent on the local rotation of the coordinate system [12]. Hence, given a deformation ϕ_n the transformed spherical function field at location x can be computed as

$$d(\mathbf{g};\mathbf{D}_n \circ \phi_n) = \sum_{i,j,k,l} D_n^{i,j,k,l}(x \circ \phi_n)(\mathbf{R}_x\mathbf{g})_i(\mathbf{R}_x\mathbf{g})_j(\mathbf{R}_x\mathbf{g})_k(\mathbf{R}_x\mathbf{g})_l \quad (4)$$

where $\mathbf{R}_x$ is the rotation of deformation ϕ_n at location x, and the notation $(\mathbf{R}_x\mathbf{g})_i$ represents the i^{th} component of the rotated vector $\mathbf{g}$.

The deformation can be parametrized as a time varying vector field such that $\partial\phi_n(x,t)/\partial t = v_n(x,t)$, $t \in [0,1]$, where $v_n(x,t)$ is the velocity field at time t. In this formulation the estimated deformation is given by $\phi_n = \phi_n(x,1) = \int_0^1 v_n(x,t)dt$. Furthermore, the $cost()$ function in Eq. 3 can be defined as $\int_0^1 \parallel Lv_n(x,t) \parallel^2 dt$, where L is a differential operator on the velocity fields [3].

We will minimize the energy function (Eq. 3) by evolving the deformation fields ϕ_n using a greedy iterative scheme which approximates the solution to the above minimization problem, similar to the technique in [3]. For this purpose we will construct a field of forces by computing the first order variation of the first term in Eq. 3 with respect to the transformation parameters as follows

$$F_n = -2 \int_{S^2} \log \left(\frac{d(\mathbf{g};\mathbf{D}_n \circ \phi_n)}{d(\mathbf{g};\mathbf{D}_\mu)} \right) [\nabla_{trans} + \nabla_{rot}] log(d(\mathbf{g};\mathbf{D}_n \circ \phi_n))d\mathbf{g} \quad (5)$$

where the variation ∇_{trans} is related with the local translation (i.e. variation of $D_n^{i,j,k,l}(x \circ \phi_n)$ in Eq. 4) and ∇_{rot} is related with the local rotation (i.e. variation of $(\mathbf{R}_x\mathbf{g})_i(\mathbf{R}_x\mathbf{g})_j(\mathbf{R}_x\mathbf{g})_k(\mathbf{R}_x\mathbf{g})_l$ in Eq. 4). The computation of these terms is discussed in Sec. 3.1.

After the estimation of the fields of forces F_n, $n = 1 \ldots N$ we compute the update vector fields $v_n = \int_\Omega K(x)F_n(x)dx$, where K is a kernel applied to the

field of forces. In our experiments we employed the kernel $K(x) = \eta(x)G(x)$, where G is a Gaussian kernel centered at x and η is a smooth function that takes zero value at the boundaries and therefore imposes zero boundary conditions on the kernel K as was done in [8]. Note that the integration of K with F_n is a convolution that becomes multiplication in the frequency domain, hence it can be efficiently computed using the discrete Fourier transform [16]. Then, the deformation fields are updated as $\phi_n^{new} = \phi_n^{old}(x + \epsilon v_n)$ using a small step ϵ.

Finally, the tensor coefficients of the atlas can be updated by also minimizing the first term in Eq. 3 with respect to the parameters of a positive definite 4^{th}-order tensor using the parametrization in [15].

3.1 Implementation Details

In general, the integral over the sphere in Eq. 5 cannot be computed analytically when the Cartesian tensor parametrization is used for modeling the diffusivity function. On the other hand the Riemannian space of ordered n-tuples (see Sec. 2) leads to analytic calculations and therefore we used it in our implementation. We constructed an m-tuple space by using a set of unit vectors $\mathbf{g}_m$ $m = 1 \dots M$ uniformly distributed on the sphere. This set of vectors can be constructed by tessellating the icosahedron and then projecting the vectors on the unit hemisphere (we consider only a hemisphere due to antipodal symmetry of diffusivity functions). We use this set of vectors in order to evaluate the spherical functions $I_{n,m} = log(d(\mathbf{g}_m; \mathbf{D}_n \circ \phi_n))$, $m = 1 \dots M$ and $n = 1 \dots N$. This creates N vector valued images I_n, whose vectors contain the M elements of the m-tuples. Note that in this m-tuple space the integrals over the sphere in Eqs. 3 and 5 become summations over m.

The above discretization helps us also in reducing the time complexity of atlas computation, which can now be efficiently computed by

$$d_\mu(\mathbf{g}_m) = exp(\frac{1}{N}\sum_{n=1}^{N} \log(d(\mathbf{g}_m; \mathbf{D}_n \circ \phi_n))) \tag{6}$$

where $d_\mu(\mathbf{g}_m)$ is also in the form of a vector valued image, whose vectors contain M elements. Note that $\log(d(\mathbf{g}_m; \mathbf{D}_n \circ \phi_n))$ is an already computed image $(I_{n,m})$, and therefore there is no need to re-deform the images and re-compute the *log* maps. The corresponding driving forces in Eq. 5 are now computed as follows

$$F_n = -2\sum_{m=1}^{M} L_{m,n}(x)\nabla I_{n,m} + \sum_{|y-x|=1} L_{m,n}(y)\frac{\nabla_{\mathbf{R}_y\mathbf{g}_m} d(\mathbf{g}_m; \mathbf{D}_n(y \circ \phi_n))}{d(\mathbf{g}_m; \mathbf{D}_n(y \circ \phi_n))}\nabla \mathbf{R}_y\mathbf{g}_m \tag{7}$$

where $L_{m,n} = \log\left(\frac{I_{n,m}(x)}{d(\mathbf{g}_m; \mathbf{D}_\mu(x \circ \phi_n))}\right)$, $\nabla\ I_{n,m}$ is simply the spatial gradient of a scalar valued image and the second term in Eq. 7 correspond to the gradient related to the tensor re-orientation. In this term the rotation $\mathbf{R}_y$ at location 'y' can be efficiently computed by the Gram-Schidt algorithm as in [8]. Using

this orthogonalization technique the components of the rotation matrix are expressed as functions of the displacement vectors in ϕ_n, hence we can easily compute analytic derivatives with the unknown transformation parameters denoted as $\nabla \mathbf{R}_y \mathbf{g}_m$. The computed derivatives are non zero for those voxels 'y' which are in the neighborhood of our current voxel 'x'. Furthermore, the gradient of the tensor with respect to the rotation is given by $\nabla_{\mathbf{R}_y \mathbf{g}_m} d(\mathbf{g}_m; \mathbf{D}_n(y \circ \phi_n)) = 4 \sum_{i,j,k} D_n^{i,j,k,l}(y \circ \phi_n)(\mathbf{R}_y \mathbf{g})_i (\mathbf{R}_y \mathbf{g})_j (\mathbf{R}_y \mathbf{g})_k$.

Finally, after the termination of the iterative minimization procedure, the 4^{th}-order tensor coefficients can be computed by fitting the tensorial model to the estimated values $d_\mu(\mathbf{g}_m)$ using the positive-definite parametrization in [15].

4 Experimental Results

In order to compare the Riemannian metric presented in Sec. 2 with a Euclidean metric in terms of fiber orientation accuracy of the atlas estimated by each metric, we performed the following experiment. We synthesized a 2-fiber crossing DW-MRI dataset (in a single voxel) using the realistic adaptive kernel model shown in Fig.3 of [17] (81 gradient directions and $b = 1250 s/mm^2$). We computed a 4^{th}-order tensor (shown in Fig. 1 upper left) from the synthetic dataset using the algorithm in [15]. Then we generated 100 more datasets by applying small rotations to the simulated crossing and by adding outliers (few of them are shown in Fig. 1 left). The computed atlases (average tensors) are compared in the bar chart of Fig. 1. As expected, the Riemannian mean outperforms the Euclidean mean since the physical space of the data is that of positive-valued functions.

To motivate the use of 4^{th}-order tensors in registering DW-MRI, we also simulated a fiber crossing dataset and synthesized a deformation field (Fig. 2). Then we computed the corresponding FA, DTI and 4^{th}-order tensor fields and their deformed images as well. We registered the obtained datasets using the scalar image registration method in [3], its DTI modification [10], and the 4^{th}-order tensor field algorithm in [12] respectively as well as our proposed method. After that, the displacement field produced by each algorithm was used to warp the deformed 4^{th}-order tensor field and it was then compared to the ground

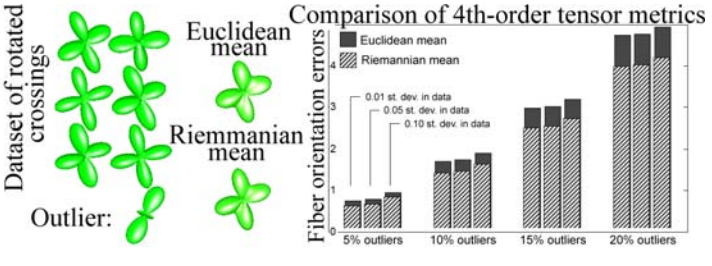


Fig. 1. Comparison of the 4^{th}-order tensor atlases computed by various metrics: a) Euclidean mean, b) Riemannian mean (computed in the space presented in Sec. 2)

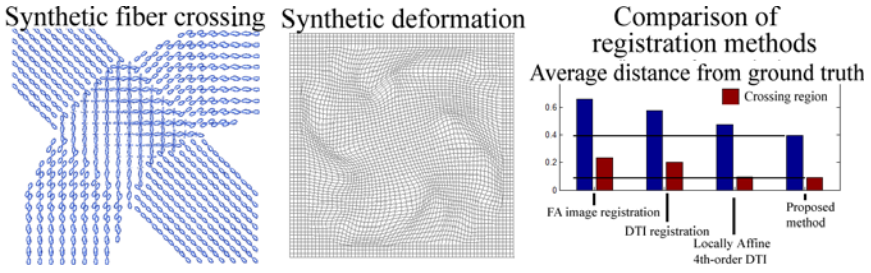


Fig. 2. Comparison of registration methods using a synthetic fiber crossing dataset. The errors were measured by evaluating Eq. 2 on the whole field (blue).

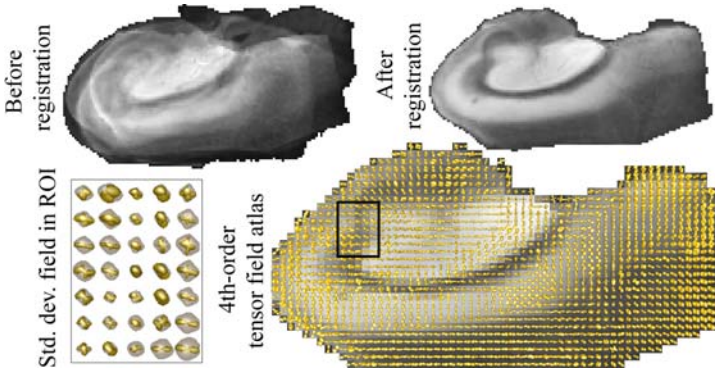


Fig. 3. Real datasets from hippocampus before and after alignment using our method. The constructed 4^{th}-order tensor field atlas is shown at the bottom. The field of standard deviations can show the variations in the dataset.

truth field shown in Fig. 2(left) using Eq. 2. The results demonstrate that our method produced more accurate mappings and registered successfully the data.

Finally, we computed the 4^{th}-order tensor field atlas from four hippocampal datasets. Each dataset consists of 21 diffusion-weighted images collected with a 415 mT/m diffusion gradient (T_d =17 ms, δ = 2.4ms, b = 1250 s/mm^2). Figure 3 shows the original misalignment of the corresponding S_0 images and the aligned images after applying our method. The 4^{th}-order tensor field atlas is depicted at the bottom of this figure and contains all the known structures of the hippocampal anatomy. The variations in the dataset can be explored by observing the standard deviation field computed by the proposed Riemannian metric (shown in an ROI on the bottom left).

5 Conclusions

In this paper we presented a novel groupwise registration and atlas construction algorithm for DWMRI data sets each of which is represented by a 4th order

tensor field. To the best of our knowledge, there is no existing literature on this topic. The key contribution of this work is the definition of a novel metric for positive valued spherical functions which was then used in the unbiased groupwise registration and atlas construction. Experimental results on comparisons with scalar and DTI registration techniques are favourable to our method.

References

1. De Craene, M., du Bois d'Aische, A., Macq, B., Warfield, S.K.: Multi-subject registration for unbiased statistical atlas construction. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 655–662. Springer, Heidelberg (2004)
2. Yeo, B.T.T., Sabuncu, M.R., Desikan, R., Fischl, B., Golland, P.: Effects of registration regularization and atlas sharpness on segmentation accuracy. Medical Image Analysis 12, 603–615 (2008)
3. Joshi, S., Davis, B., Jomier, A., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)
4. Park, H., Bland, P.H., Hero III, A.O., Meyer, C.R.: Least biased target selection in probabilistic atlas construction. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 419–426. Springer, Heidelberg (2005)
5. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. TMI 20(11), 1131–1139 (2001)
6. Ruiz-Azola, J., Westin, C.F., Warfield, S.K., Alberola, C., Maier, S., Kikinis, R.: Non rigid registration of 3d tensor medical data. Med. Im. Anal. 6, 143–161 (2002)
7. Zhang, H., Yushkevich, P., Gee, J.: Registration of DTI. In: CVPR, vol. 1, pp. 842–847 (2004)
8. Cao, Y., Miller, M., Mori, S., Winslow, R., Younes, L.: Diffeomorphic matching of diffusion tensor images. In: Computer Vision and Pattern Recognition Workshop, June 2006, p. 67 (2006)
9. Yeo, B., Vercauteren, T., Fillard, P., Pennec, X., Gotland, P., Ayache, N., Clatz, O.: DTI registration with exact finite-strain differential. In: ISBI, pp. 700–703 (2008)
10. Zhang, H., Yushkevich, P.A., Rueckert, D., Gee, J.C.: Unbiased white matter atlas construction using diffusion tensor images. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 211–218. Springer, Heidelberg (2007)
11. Irfanoglu, M.O., Machiraju, R., Sammet, S., Pierpaoli, C., Knopp, M.V.: Automatic deformable diffusion tensor registration for fiber population analysis. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 1014–1022. Springer, Heidelberg (2008)
12. Barmpoutis, A., Vemuri, B.C., Forder, J.R.: Registration of high angular resolution diffusion mri images using 4th order tensors. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 908–915. Springer, Heidelberg (2007)
13. Özarslan, E., Mareci, T.H.: Generalized diffusion tensor imaging and analytical relationships between dti and hardi. MRM 50(5), 955–965 (2003)
14. Jensen, J.H., Helpern, J.A., Ramani, A., Lu, H., Kaczynski, K.: DKI:the quantification of non-gaussian water diffusion by means of MRI. MRM 53(6), 1432–1440 (2005)
15. Barmpoutis, A., Hwang, M.S., Howland, D., Forder, J.R., Vemuri, B.C.: Regularized positive-definite fourth order tensor field estimation from DW-MRI. NeuroImage 45(1 supl. 1), 153–162 (2009)
16. Joshi, S., Grenander, U., Miller, M.I.: On the geometry and shape of brain submanifolds. IJPRAI 11(8), 1317–1343 (1997)
17. Barmpoutis, A., Jian, B., Vemuri, B.C.: Adaptive kernels for multi-fiber reconstruction. In: IPMI 2009. LNCS, vol. 5636, pp. 338–349. Springer, Heidelberg (2009)

Closed-Form Jensen-Renyi Divergence for Mixture of Gaussians and Applications to Group-Wise Shape Registration*

Fei Wang[1], Tanveer Syeda-Mahmood[1], Baba C. Vemuri[2], David Beymer[1], and Anand Rangarajan[2]

[1] IBM Almaden Research Center, San Jose, CA, USA
[2] Department of CISE, University of Florida, Gainesville, FL, USA

Abstract. In this paper, we propose a generalized group-wise non-rigid registration strategy for multiple unlabeled point-sets of unequal cardinality, with no bias toward any of the given point-sets. To quantify the divergence between the probability distributions – specifically Mixture of Gaussians – estimated from the given point sets, we use a recently developed information-theoretic measure called Jensen-Renyi (JR) divergence. We evaluate a closed-form JR divergence between multiple probabilistic representations for the general case where the mixture models differ in variance and the number of components. We derive the analytic gradient of the divergence measure with respect to the non-rigid registration parameters, and apply it to numerical optimization of the group-wise registration, leading to a computationally efficient and accurate algorithm. We validate our approach on synthetic data, and evaluate it on 3D cardiac shapes.

1 Introduction

The non-rigid registration of multiple point sets is a fundamental problem in medical imaging for atlas generation and disease class representation. In model-based image segmentation for example, constructing an atlas typically requires us to bring pre-segmented shapes into alignment. Similarly, registration of hippocampal regions across patients with a given disease helps capture the essential disease-specific variations of the region. This is particularly relevant for echocardiography where a disease-specific understanding of echocardiagraphic sequences requires accurate characterization of spatio-temporal motion patterns in the echo videos. The supplemental videos show different patients diagnosed with the same disease of hypokinesia. The similar motion patterns in the designated cardiac regions can be observed. Simultaneous registration of such videos helps summarize the motion pattern for a disease class.

Simultaneous registration of such shapes can be a daunting task due to several reasons. First, the shapes are deformable requiring non-rigid spatial transformations. Secondly, explicit correspondence through search is computationally

* This research was in part funded by the NIH grant RO1-NS046812 and NSF grant NSF 0307712.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 648–655, 2009.

prohibitive for such large data sets. Robustness to missing and spurious features would be desired. Finally, the registration must be *unbiased* towards any one of the shapes.

In this paper, we develop a novel non-rigid registration technique for multiple point-sets based on a recently introduced information theoretic matching criterion [1] called Jensen-Renyi (JR) divergence. We show that using Renyi's Quadratic entropy results in a closed-form solution for the registration problem. We also derive the analytic gradient of this match measure in order to achieve efficient and accurate non-rigid registration. The Jensen-Renyi measure is then minimized over a class of smooth non-rigid transformations expressed in a thin-plate spline basis. We show the results of our algorithm for the registration of medical imaging datasets from spatio-temporal cardiac echo videos.

2 Previous Work

The work reported here fits into the general class of approaches that avoid explicit point correspondences for non-rigid registration through the use of divergence measures between probability distributions formed around point sets [2,3,4]. Specifically, the Jensen-Shannon (JS) divergence and the CDF-based Jensen-Shannon divergence between the feature point-sets was minimized to obtain the non-rigid deformation. The density-based approaches are relatively more robust to the point-sets of different sizes and to the presence of missing features. Furthermore, if an unbiased information theoretic measure is chosen to quantify the multiple densities representing the shapes, the matching results can potentially be unbiased to any of the given point-sets [4].

Other related work includes Wang et al. [5], where the relative entropy measure (Kullback-Leibler divergence) is used to find a similarity transformation between two point-sets under rigid transformation. In Jian et al. [2] and Roy et al. [6], nonrigid registration is between pairs of data sets using L2 distance on a mixture of Gaussians model of the data sets. Both methods, however, have not been extended to the problem of unbiased simultaneous matching of multiple point-sets being addressed in this paper. Recently, in [7], Glaunes et al. represent points as delta functions and match them using the dual norm in a reproducing kernel Hilbert space. The resulting 3D spatial integral is difficult to compute numerically. Similarly, the JS divergence proposed earlier [3] was estimated using the law of large numbers, making it computationally expensive.

In [1], a new divergence measure called the Jensen-Renyi divergence was introduced as a generalization of the Jensen-Shannon divergence being based on Renyi Entropy [1]. In contrast to the JS divergence used earlier [3], the exponential order of Renyi entropy as well as the weights to control the measurement sensitivity of the probability densities can be adjusted in JR divergence. Further, the JR divergence measure enjoys appealing mathematical properties such as convexity and symmetry, affording a great flexibility in a number of applications [1]. Finally, a related work by Chiang et al. [8] has used JR divergence for pairwise image matching where contiguity constraints between pixels could be utilized.

Our approach differs from prior work in many respects. First, we consider a new problem of simultaneous registration of multiple points sets (rather than multiple pair-wise registrations). Secondly, we introduce the JR divergence for the point-sets registration problem and give a closed-form solution to divergence estimation making it computationally attractive. More importantly, Jensen-Renyi divergence can be generalized to a family of divergence measures among a group of point-sets, with JS-divergence being a special case (when $\alpha \rightarrow 1$). Finally, our method can accommodate point-sets to be registered of varying size, and is also robust in the presence of noise.

3 Multiple Point-Sets Registration Using JR Divergence

3.1 Shape Representation by Mixture of Gaussians

Consider first the representation of probability density of shapes as Mixture of Gaussian (MoG). Let the N point-sets to be registered be denoted by $\{X^{(i)}, i \in \{1, ..., N\}\}$. Each point-set $X^{(i)}$ consists of d-dimensional points $\{x_j^{(i)} \in \mathbb{R}^d, j \in \{1, ..., n_p\}\}$. Each point set is represented by a probability density function P_i. The features of i^{th} shape $X^{(i)}$ are represented as a Gaussian Mixture Model

$$\begin{aligned} P_i = p(\mathbf{x}|\theta^{(i)}) &= \sum_{a=1}^{K_i} \omega_a^{(i)} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma_a^{(i)}|^{\frac{1}{2}}} \exp\Big(-\frac{1}{2}\big(\mathbf{x} - u_a^{(i)}\big)^T (\Sigma_a^{(i)})^{-1} \big(\mathbf{x} - u_a^{(i)}\big)\Big) \\ &= \sum_{a=1}^{K_i} \omega_a^{(i)} G(\mathbf{x} - u_a^{(i)}, \Sigma_a^{(i)}) \end{aligned} \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^d$, parameter set $\theta^{(i)} = \{\omega_a^{(i)}, u_a^{(i)}, \Sigma_a^{(i)}\}$, and $G(\mathbf{x} - u_a^{(i)}, \Sigma_a^{(i)})$ is the Gaussian Kernel in d-dimensional space. Constraints on the MoG weight $\omega_a^{(i)}$ are $\{\omega_a^{(i)} > 0, \sum_{a=1}^{K_i} \omega_a^{(i)} = 1\}$. Since $X_j^{(i)}$ are assumed to be i.i.d., the likelihood of the set of features of $X^{(i)}$ is

$$p(X^{(i)}|\theta^{(i)}) = \prod_{j=1}^{N_i} \sum_{a=1}^{K_i} \omega_a^{(i)} G(X_j^{(i)} - u_a^{(i)}, \Sigma_a^{(i)}) \tag{2}$$

For all the shapes, we fit the model parameters $\theta^{(i)}$s by minimizing the negative log-likelihood objective function in Eqn.(2) w.r.t. the model parameters. For simplicity, we specialize to the case where the occupancy probabilities are uniform $\omega_a^{(i)} = \frac{1}{K_i}$ and we have isotropic covariance matrix $\Sigma_a^{(i)} = \sigma_i^2 \mathbf{I}_d$ for all the shapes.

The minimization of Eqn.(2) can be easily accomplished using the well-known EM algorithm for the MoG [9]. This computation is done offline for all the shapes once we fix the number of centroids K_i. Model selection for mixture models needs to be performed to fix the number of centroids.

3.2 A Closed-Form Divergence Measure between Multiple MoGs

We now derive the cohesive measure between the multiple probabilistic shape representations with the help of the Jensen-Renyi divergence. Please note that distance measures like Jensen-Shannon [3] cannot be derived in closed form for the Gaussian mixture model.

The Jensen-Renyi divergence [1] between probability density functions is defined as:

$$JR_\pi(P_1, P_2, ..., P_N) = H_\alpha(\sum \pi_i P_i) - \sum \pi_i H_\alpha(P_i), \tag{3}$$

where $\pi = \{\pi_1, \pi_2, ..., \pi_n | \pi_i > 0, \sum \pi_i = 1\}$ are the weights of the probability distributions P_i, $H_\alpha(X)$ is Renyi entropy of order α, $\alpha \geq 0$ defined on a random variable X and is given by $H_\alpha(X) = \frac{1}{1-\alpha} \log(\sum_{i=1}^{n} p_i^\alpha)$, where p_i are the probabilities of $x_1, x_2, ..., x_n$. Note that Renyi entropy is a generalization of Shannon entropy, since $\lim_{\alpha->1} H_\alpha = H_S$, and H_S is Shannon entropy of the same random variable. Thus Jensen-Shannon divergence is a special case of Jensen-Renyi divergence when $\alpha \to 1$. In the definition of the α-Renyi Entropy, when $\alpha = 2$, $H_2 = -\log(\sum_{i=1}^{n} p_i^2)$ is called quadratic entropy. We now show that in the case of quadratic entropy ($\alpha = 2$), there exists a closed-form solution to the estimation of JR divergence between multiple MoGs, i.e. $JR(p(\mathbf{x}|\theta^{(1)}), \ldots, p(\mathbf{x}|\theta^{(N)}))$.

Given a Gaussian mixture model $\hat{p}_V(x) = \sum_{i=1}^{N} \omega_i G(x - v_i, \sigma^2 \mathbf{I})$, where v_i and $\sigma^2\mathbf{I}$ are the centroids and covariance matrix respectively, unlike Shannon entropy, Renyi's quadratic entropy has a closed form for MoGs

$$\begin{aligned} H_2(V) = -\log \int_{-\infty}^{+\infty} \hat{p}_V(x)^2 dx = -\log \Big[\sum_{j=1}^{N} \int_{-\infty}^{+\infty} \omega_i \omega_j G(x - v_i, \sigma^2 \mathbf{I}) \\ G(x - v_j, \sigma^2 \mathbf{I}) dx \Big] = -\log \Big[\sum_{i=1}^{N} \sum_{j=1}^{N} \omega_i \omega_j G(v_i - v_j, 2\sigma^2 \mathbf{I}) \Big], \end{aligned} \tag{4}$$

where $v_i \in \mathbb{R}^d$ and $v_j \in \mathbb{R}^d$ are two centroids, and Σ_i and Σ_j are two covariance matrices for two Gaussian kernels in the space. Substituting the value of quadratic entropy in Eqn. (3) for the MoG of each shape expressed in Eqn. (1), we get an estimate for the quadratic entropy of each individual probabilistic shape representation

$$H_2(P_i) = -\log \Big[\frac{1}{K_i^2} \sum_{k=1}^{K_i} \sum_{l=1}^{K_i} G(u_k^{(i)} - u_l^{(i)}, 2\sigma_i^2 \mathbf{I}) \Big]. \tag{5}$$

For the convex combination $\sum \pi_i P_i$, if we choose $\pi_i = \frac{K_i}{M}$, where $M = \sum_i K_i$ is the total number of the centroids in all shapes, we have the following,

$$\sum_{i=1}^{N} \pi_i P_i = \frac{1}{M} \sum_{i=1}^{N} \sum_{a=1}^{K_i} G(\mathbf{x} - u_a^{(i)}), \sigma_i^2 \mathbf{I}) = \frac{1}{M} \sum_{j=1}^{M} G(\mathbf{x} - u_j, \sigma_{\tau(j)}^2 \mathbf{I}) \tag{6}$$

where $\{u_1, u_2, ..., u_M\} \equiv \{u_1^1, ..., u_j^i, ..., u_{K_N}^N\}$ are the pooled centroids, $\tau : \{1, ..., M\} \to \{1, ..., N\}$ is a mapping function that maps the index of an individual point to the index of the point-set. Therefore the linear combination

of the MoGs can be expressed as a single Gaussian Mixture centered on the pooled centroids. Consequently, we have the quadratic entropy estimation of the $\sum \pi_i P_i$,

$$H_2(\sum_{i=1}^{N} \pi_i P_i) = -\log\Big[\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} G(u_i - u_j, (\sigma^2_{\tau(i)} + \sigma^2_{\tau(j)})\mathbf{I})\Big] \tag{7}$$

Combining the two terms in Eqn.(5) and (7) together, we have the closed-form JR-divergence for multiple MoGs,

$$\begin{aligned} JR_\pi(P_1, P_2, ..., P_N) = H_2(\sum \pi_i P_i) - \sum \pi_i H_2(P_i) = -\log\Big[\frac{1}{M^2}\sum_{i=1}^{M}\sum_{j=1}^{M} \\ G(u_i - u_j, (\sigma^2_{\tau(i)} + \sigma^2_{\tau(j)})\mathbf{I})\Big] + \sum_{i=1}^{N}\frac{K_i}{M}\log\Big[\frac{1}{K_i^2}\sum_{k=1}^{K_i}\sum_{l=1}^{K_i} G(u^i_k - u^i_l, 2\sigma_i^2\mathbf{I})\Big] \end{aligned} \tag{8}$$

The JR divergence can be further simplified using matrix notation. Let us introduce a Gaussian kernel matrix $\mathbf{G}$ with $G_{ij} = G(u_i - u_j, (\sigma^2_{\tau(i)} + \sigma^2_{\tau(j)})\mathbf{I})$, and define an indicator vectors I_a (of length M) for a^{th} point-set, i.e. $I_a(i) = 1$ if u_i is from the a^{th} point-set, i.e. $\tau(i) = a$. I_M is the vector whose elements are all ones. Eqn. (8) can be rewritten as

$$JR_\pi(P_1, P_2, ..., P_N) = -\log\frac{I_M^T\mathbf{G}I_M}{M^2} + \sum_i \frac{K_i}{M}\log\frac{I_i^T\mathbf{G}I_i}{K_i^2} \tag{9}$$

This is the final expression for the distance function used in this paper. The number of centroids and the variances of each shape can be different.

3.3 Cost Function Optimization

We now turn to the description of the deformation model. We assume that parameters of each shape $\theta^{(i)} = \{u^{(i)}, \sigma_i\}$ (comprising the centroids and variance) are simultaneously deforming until the JR divergence between $\{p(\mathbf{x}|\theta^{(i)}), i \in \{1, \ldots, N\}\}$ are minimized. We use the familiar thin-plate spline (TPS) deformation model for the centroids, for which we refer the readers to [3] for more details. In addition a regularization term is added to JR divergence to prevent reflections and unphysical affine transformations.

Based on Eqn.(9), we can derive the gradient of the JR divergence with respect to the transformation parameters $\boldsymbol{\mu}^{(a)}$, which is given by,

$$\frac{\partial JR}{\partial \boldsymbol{\mu}^a} = -\frac{I_M^T\frac{\partial \mathbf{G}}{\partial \boldsymbol{\mu}^a}I_M}{I_M^T\mathbf{G}I_M} + \sum_{i=1}^{N}\frac{K_i}{M}\frac{I_i^T\frac{\partial \mathbf{G}}{\partial \boldsymbol{\mu}^a}I_i}{I_i^T\mathbf{G}I_i} \tag{10}$$

The details of the derivation are omitted due to space limitations. Once we have the analytical gradient, the cost function optimization is achieved very efficiently using a quasi-Newton method. The variance parameters σ_i are updated separately from the remaining parameters and are not allowed to abruptly change.

4 Results

We now present experimental results on both synthetic and real data sets. We first demonstrate the robustness and accuracy that can be achieved in registration with JR divergence in comparison to Jensen-Shannon divergence reported in an earlier work. We then report results of registration on data sets drawing from cardiac imaging.

4.1 Robustness to Noise and Outliers

Our approach to validation is based on comparing the recovered transformation parameters against the true (synthetically generated) parameters for a randomly generated 2D point-set $\mathcal{M}$ (consisting of 300 points). A rigid transformation is applied to the point-set $\mathcal{M}$ to obtain a transformed point-set $\mathcal{S}$, and we then remove 50% of the points in $\mathcal{S}$ to get a reduced set and this is done so that the two mixture densities have a large discrepancy in the number of centroids. We then execute the registration at 30 different noise trials for each different rigid transformation. For each pair of point-sets, we use our algorithm and the Jensen-Shannon method to estimate the known rigid transformation which was partially responsible for the corruption. Results show that our method exhibits stronger resistance to noise than the JS method, as shown in Table 1. Furthermore, the average running time for all synthetic examples is $2.39s$ in our method compared with $3.96s$ in the JS algorithm (single core, 1.66 GHz processor, both algorithms are implemented in Matlab).

Table 1. Robustness test in the presence of large noise

JR improvement over JS in $\mathbf{R}, \mathbf{t}$ estimation (%)	outlier fraction $(\tau - \rho)/(\rho)$				
	0.1	0.2	0.3	0.4	0.5
rotation $\mathbf{R}$ improvement (%)	11.23%	13.57%	4.36%	27.91%	14.90%
translation $\mathbf{t}$ improvement (%)	5.26%	7.24%	3.55%	-3.53%	13.83%

4.2 Evaluation on 3D Shapes from Cardiac Echo Videos

Next, we presents results over 3D shapes extracted from cardiac echo videos. Specifically, cardiologists from hospitals in India provided us echo videos of patients with various diseases, which depicted over 500 heart beat cycles chosen from over 50 patients with a number of cardiac diseases including myocardial infarction, cardiomyopathy, mitral regurgitation, and regional wall motion abnormality etc. For each disease class, we collected videos depicting similar views (short axis, long axis, four chamber views). An Active Shape Model (ASM) was used to characterize each such view as described in [10]. Feature points corresponding to identifiable landmarks on heart wall boundaries were automatically extracted and tracked [10] to obtain a 3D point set.

We showcase results on 3D shapes of Parasternal Long Axis (PLA) views from five different patients, which are formed by stacking together the points from all frames of a single cycle, and the time axis is normalized for each echocardiogram

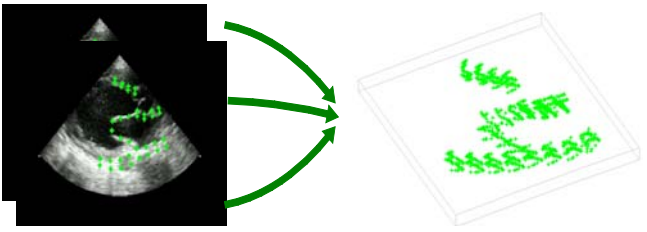


Fig. 1. Illustration of echo feature points extracted from echo video of the Parasternal Long Axis (PLA) view

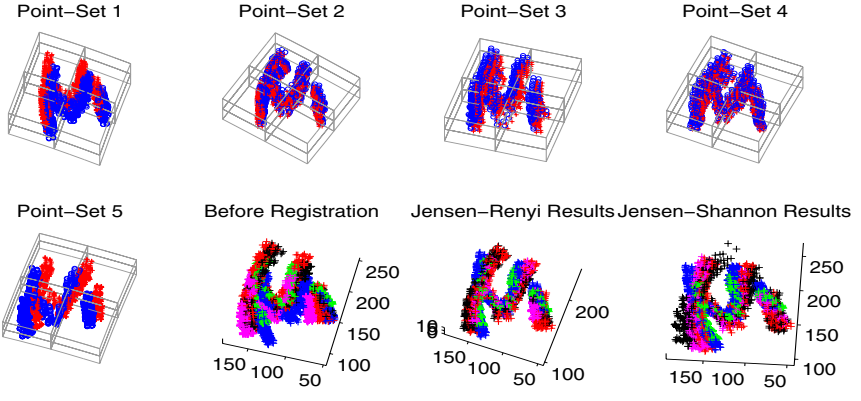


Fig. 2. Experiment results on five 3D echocardiogram point-sets. (see text for details).

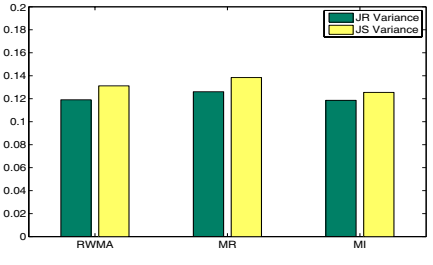


Fig. 3. Within different cardiac diseases, registration variance is lower with our JR approach compared with JS, showing we achieve a better overall registration

video with respect to the heart beat of the patient (Fig. 1). As shown in Fig. 2, the recovered deformation between each point-set (blue) and the mean shape (red) are superimposed on the first row and the left image in second row in Figure 2. The scatter plot of the five point-sets before and after registration results are shown in second and third image of the second row of Figure 2. The registration results generated using Jensen-Shannon is shown in the lower-right for comparison, from which we can observe that the results generated using our algorithm exhibits more similarity than the JS approach. As we described earlier, all these results are

computed simultaneously and automatically. This example demonstrates that our matching algorithm can simultaneously align multiple shapes even though the information on point correspondences is not in anyway available.

The advantage of JR over JS in registering point sets also exists within disease categories. In Fig. 3, we show the remaining point variance after registration of videos from a number of diseases: regional wall motion abnormality (RWMA), mitral regurgitation (MR), and myocardial infarction (MI). For all diseases, the JR variance is lower than the corresponding JS variance, showing that JR performs a superior registration of the point sets.

5 Conclusions

In this paper, we presented a novel and robust algorithm that utilizes an information theoretic measure, namely Jensen-Renyi divergence, to simultaneously register multiple unlabeled point-sets. Using the quadratic form of Renyi entropy, we have shown that it is possible to obtain a closed-form solution to the non-rigid registration problem leading to computational efficiency in registration. While we used Gaussian kernels to represent the probability density of point sets, the formalism holds for other kernels as well. Experiments were depicted with both 2D and 3D point sets from medical and non-medical domains. Future work will focus on generalizing the non-rigid deformations to diffeomorphic mappings.

References

1. Hamza, A.B., Krim, H.: Jensen-renyi divergence measure: theoretical and computational perspectives. In: ISIT, p. 257 (2003)
2. Jian, B., Vemuri, B.: A robust algorithm for point set registration using mixture of gaussians. In: ICCV 2005, pp. 1246–1251 (2005)
3. Wang, F., Vemuri, B.C., Rangarajan, A., Eisenschenk, S.J.: Simultaneous nonrigid registration of multiple point sets and atlas construction. In: PAMI, pp. 2011–2022 (2008)
4. Wang, F., Vemuri, B.C., Rangarajan, A.: Groupwise point pattern registration using a novel CDF-based jensen-shannon divergence. In: CVPR, pp. 1283–1288 (2006)
5. Wang, Y., Woods, K., McClain, M.: Information-theoretic matching of two point sets. IEEE Transactions on Image Processing 11(8), 868–872 (2002)
6. Roy, A.S., Gopinath, A., Rangarajan, A.: Deformable density matching for 3D non-rigid registration of shapes. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 942–949. Springer, Heidelberg (2007)
7. Glaunes, J., Trouvé, A., Younes, L.: Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching. In: CVPR, pp. 712–718 (2004)
8. Chiang, M.C., Dutton, R.A., Hayashi, K.M., Toga, A.W., Lopez, O.L., Aizenstein, H., Becker, J.T., Thompson, P.M.: Fluid registration of medical images using jensen-renyi divergence reveals 3d profile of brain atrophy in HIV/AIDS. In: ISBI, pp. 193–196 (2006)
9. Yuille, A.L., Stolorz, P., Utans, J.: Statistical physics, mixtures of distributions, and the EM algorithm. Neural Comput. 6(2), 334–340 (1994)
10. Syeda-Mahmood, T., Wang, F., Beymer, D., London, M., Reddy, R.: Characterizing spatio-temporal patterns for disease discrimination in cardiac echo videos. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 261–269. Springer, Heidelberg (2007)

Attribute Vector Guided Groupwise Registration

Qian Wang[1,2], Pew-Thian Yap[2], Guorong Wu[2], and Dinggang Shen[2,*]

[1] Department of Computer Science, University of North Carolina at Chapel Hill
qianwang@cs.unc.edu
[2] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
{ptyap,grwu,dgshen}@med.unc.edu

Abstract. Groupwise registration has been recently introduced for simultaneous registration of a group of images with the goal of constructing an unbiased atlas. To this end, direct application of information-theoretic entropy measures on image intensity has achieved various successes. However, simplistic voxelwise utilization of image intensity often neglects important contextual information, which can be provided by more comprehensive geometric and statistical features. In this paper, we employ attribute vectors, instead of image intensities, to guide groupwise registration. In particular, for each voxel, the attribute vector is computed from its multiple-scale neighborhoods to capture geometric information at different scales. Moreover, the probability density function (PDF) of each attribute in the vector is then estimated from the local neighborhood, providing a statistical summary of the underlying anatomical structure. For the purpose of registration, Jensen-Shannon (JS) divergence is used to measure the PDF dissimilarity of each attribute at corresponding locations of different individual images. By minimizing the overall JS divergence in the whole image space and estimating the deformation field of each image simultaneously, we can eventually register all images and build an unbiased atlas. Experimental results indicate that our method yields better registration quality, compared with a popular groupwise registration method.

1 Introduction

Image registration has been one of the focuses in medical imaging research for decades. It is a key step for many medical image analysis based applications, since inter-subject comparison cannot be performed without first normalizing all images onto a common space. To this end, though a lot of methods have been reported, most of them are, however, pairwise in nature, where a floating image is warped to the space of a fixed image. When pairwise approaches are directly applied to a population of images, one subject needs to be selected in advance as the fixed image, to which all other images can be registered. This manual determination inevitably biases the registration and also subsequent analysis. In light of this, recently proposed groupwise registration algorithms consider the image population as a whole, by simultaneously deforming all individual images impartially, with the final goal of constructing an atlas which describes the population in an unbiased manner [1][2]. Inferences drawn

* Corresponding author.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 656–663, 2009.

using an unbiased atlas can be expected to be more objective, and reflect more accurately the population characteristics.

Several groupwise registration methods have been proposed in literature. Joshi *et al.* [3] extends a large deformation diffeomorphic mapping algorithm [4] to work in a groupwise manner for unbiased atlas construction. A congealing framework [5] based groupwise registration scheme is proposed in [6], where intensity based entropy drives a gradient-based stochastic optimizer and pushes each image to the population center simultaneously. This method, which originally works for affine transformation only, is further extended by Balci *et al.* [7] to incorporate B-Splines to model non-rigid deformation. The objective function, called stack entropy in [6] and [7], measures the compactness of the intensity distribution for a certain voxel location across different subjects. Similar to the popular entropy based metrics used in pairwise registration (e.g., mutual information [8][9]), the current formulation of stack entropy considers image intensity as the only feature, and discards local contextual information that can be provided by the voxel neighborhood. Moreover, each voxel contributes equally to the metric, regardless of its anatomical properties.

In this paper, we propose a novel groupwise registration algorithm, which utilizes multi-scale attribute vectors, rather than only image intensities, as feature descriptors. To better capture the variability of each attribute in the attribute vector, its probability density function (PDF) is further estimated from the neighborhood. Then, the registration algorithm aims to minimize the overall PDF divergence of attribute vectors at each corresponding location across different images. Finally, all the images are groupwisely registered onto a common group mean. Experimental results indicate that our method yields better registration quality and outperforms Balci's method [7].

2 Method

Multi-scale attribute vectors are used as voxel anatomical signatures to guide groupwise registration. For a given voxel location, the probability density function (PDF) of each attribute in the vector is estimated from the neighborhood of the voxel. The objective of our groupwise registration is then to minimize the overall PDF divergence of the whole attribute vector, across different subjects, and for each voxel location. The final result is the registration of all images onto a common space defined by the estimated atlas. We will describe our method in more details in the following sections.

2.1 Attribute Vectors and Their Statistical Descriptions

In medical image analysis, it is important to build anatomical models that take into account the underlying anatomy, rather than only the similarity of image intensities. Moreover, brain images have intrinsically high anatomical variation, and hence more complex descriptors are needed to fully exploit the structural information. This motivates the utilization of more descriptive features, which take into consideration the different shapes and sizes of anatomical details.

Attribute vectors can capture rich anatomical information, and have been successfully applied to achieve accurate pairwise registration [10]. The composition of the

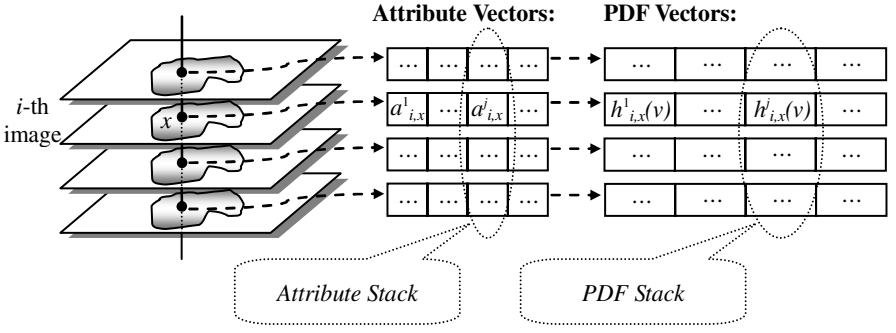


Fig. 1. Attribute vectors at the voxel location. Attribute vectors at the same corresponding location across all subjects form an attribute stack. PDF vectors and PDF stacks are further produced, providing statistical descriptions of attributes in the neighborhood of voxel.

attribute vector is quite flexible and can be easily fine-tuned to cater for different applications. As shown in Fig. 1, we formulate an attribute vector at voxel location x of the i-th image as $\vec{A}_{i,x} \triangleq \{a^j{}_{i,x} | j = 1{:}\, D\}$, where $\boldsymbol{D}$ denotes the number of attributes in the vector. Each attribute $a^j{}_{i,x}$ is a scalar and describes a specific feature, which can constitute the voxel intensity, and also more contextual features such as gradient and directional derivatives. It is, however, worth noting that if the attribute is derived from only a single voxel, the attribute vector can be susceptible to imaging noise and anatomical ambiguity. A better approach is to include attributes which describe contextual information of each voxel by considering its neighborhood. Attributes as such are more robust to imaging noise and can be expected to improve the robustness of the registration. Attributes at the same location across all the subjects then form an attribute stack (as shown in Fig. 1), and we note here that the attribute vectors are always compared stack-wise.

The optimal composition of attribute vectors is related to the specific application. Though complicated features could bring potential benefits to groupwise registration, simple attributes can greatly save computational time, which is critical to groupwise registration where many images need to be handled simultaneously. Therefore, in our study here, only intensities and gradients (calculated in multiple-scale neighborhoods) are used to form the attribute vector, though more complex attributes (i.e., moments [10]) can be incorporated in future. It is demonstrated in the following experiments that sufficiently good performance can be achieved via simple attribute vectors.

To increase the robustness of groupwise registration, a statistical description, in the form of PDF, is estimated for each voxel by considering its neighborhood. And for each attribute vector, we define its corresponding PDF vector. A simple but effective means of estimating PDF $h^j{}_{i,x}(v)$, where v is a random variable, for the attribute $a^j{}_{i,x}$, is by evaluating the histograms of neighboring attribute values around the voxel location x. A Gaussian kernel is then used to smooth the histogram. With the PDFs estimated for all attributes in the attribute vector, we can define the PDF vector as $\vec{P}_{i,x} \triangleq \{h^j{}_{i,x}(v) | j = 1{:}\, D\}$, elements of which now replace roles of scalar attributes $a^j{}_{i,x}$ in the groupwise registration.

2.2 Jensen-Shannon Divergence and PDF Stack Divergence

Jensen-Shannon (JS) divergence [11] is an information-theoretic quantity for measuring the dissimilarity or the distance of two probability distributions. JS divergence was later adopted in [12] for groupwise registration of different point sets. The approach taken was based on the global PDF of each point set. We take a different approach by applying JS divergence on local PDFs, instead on a single global distribution. Noting that the PDF stack for the *j*-th attribute at the location x of N different images is $\{h^j{}_{i,x}(v)|i = 1{:}\,N\}$, then, the local JS divergence for the PDF stack, or *PDF stack divergence*, can be written as:

$$JS^j{}_x = \mathcal{H}_p\left(\frac{1}{N}\sum_{i=1}^{N} h^j{}_{i,x}(v)\right) - \frac{1}{N}\sum_{i=1}^{N}\mathcal{H}_p\left(h^j{}_{i,x}(v)\right) \tag{1}$$

where $\mathcal{H}_p(\cdot)$ is the Shannon entropy with respect to random variable v, and each individual image is weighted with equal prior $(1/N)$. By rewriting (1) as:

$$JS^j{}_x = \frac{1}{N}\sum_{i=1}^{N}\int h^j{}_{i,x}(v)\cdot\log\frac{h^j{}_{i,x}(v)}{\sum_{s=1}^{N} h^j{}_{s,x}(v)/N}\,dv \tag{2}$$

we can observe that the local JS divergence for a specific PDF stack measures the variation of the PDFs from the mean PDF, obtained by averaging the PDFs across N different subjects.

2.3 Overall JS Divergence

We then formulate the overall JS divergence based on all local PDF stack divergences. It is worth noting that different attributes in the vector are usually orthogonal (e.g. different directional derivatives) or statistically independent (e.g. intensity and gradient). To ensure measurements of local PDF stack divergences at different locations are minimally correlated, the locations for consideration are randomly and sparsely sampled from the image space. Therefore, the overall JS divergence, as the objective function of our groupwise registration algorithm, can be represented by the linear combination of PDF stack divergences of all attributes and all voxel locations in the image volume, that is:

$$JS_{overall} = \sum_x\sum_j w_j \cdot JS^j{}_x \tag{3}$$

where w_j is the empirically determined weighting factor for the *j*-th attribute and complies to $\sum_j w_j = 1$. Suppose $d_{i,x}$ represents the new location transformed from x for the *i*-th image ($d_{i,x} = x + \Delta x_i$), groupwise registration is achieved by minimizing the overall JS divergence in (3) through a gradient based optimization approach, following:

$$\frac{\partial JS_{overall}}{\partial d_{i,x}} = \sum_x\sum_j w_j \cdot \frac{\partial JS^j{}_x}{\partial d_{i,x}} \tag{4}$$

Generally, the optimization framework similar to [7] can be adopted. The B-Spline deformation model is employed in our method to represent the spatial deformation fields $\{d_{i,x}\}$. The optimizer will simultaneously estimate individual transformations for all images, through minimizing Equation (3). The mean identity transformation constraint is enforced to solve the global drift issue. To avoid local minima, which are quite prone to happen in a high degrees-of-freedom (DOF) optimization problem, a multi-resolution strategy is employed. The attribute vectors and their corresponding PDF vectors are computed at different scales. At a coarser level, the optimization problem is effectively reduced to a lower DOF version, and many local minima can be avoided. The registration is progressively refined as the finer resolutions are considered.

3 Experimental Results

We have employed the NA0 Database [13] to demonstrate the capability of our method. In the dataset, there are 16 individual brain MR images, each of which has 32 manual anatomical labels. Balci's method [7] is used as a benchmark since it is open-source and is freely available in ITK. Also, it is worth mentioning that our method is implemented by modifying Balci's method in ITK, so that the improvement yielded by the utilization of attribute vectors can be quantified. For fair comparison, we use the same multi-resolution registration strategy (three resolutions) and the same configuration (DOF of transformations and number of iterations) for both Balci's method and our method. After groupwise registration, all aligned images are averaged voxelwise to form the group mean which defines a common space.

To quantify the performance, we calculate the intensity residual errors between 16 aligned images and their group mean. Then, an averaged standard deviation volume for the 16 images is produced. In Fig. 2 (a) and (b), same slices of the residual error volumes are shown at different resolutions for visual comparison. Results yielded by our method are generally darker, especially in areas marked by the ovals, implying lower residual errors. It is worth noting that higher residual errors are more likely to happen near the cortical region, where the inter-subject anatomical variability is generally higher. Lower residual errors usually mean that images are better aligned, and imply higher registration accuracy. The distributions of the residual errors are also shown in Fig. 2 (c). Unsurprisingly, our method yields a distribution with peaks moved closer to zero. In fact, the average of residual errors provided by our method is 3.56% lower than that of Balci's method. Given that the same number of iterations and similar computation time is used, the improvement brought about by our method in terms of residual errors is attractive.

The ability to establish correspondences of the same anatomical structures from different brains is essential for consistent groupwise registration. We can quantitatively measure this by utilizing the estimated transformation for each image to warp the manually delineated anatomical labels to the common space, and then compare the overlap rates of these labels. The overlap rate utilized here is the Jaccard coefficient, which is defined as the ratio between the size of the region of intersection to the size of the regional union of a registered image and the mean image. In Fig. 3, we provide the average overlap rates for all 32 labels, produced via Balci's method and our method, respectively. Our method shows improvement in all 32 labels, with an average of 2.45% increase, implying that more consistent correspondence matching can be achieved using our method.

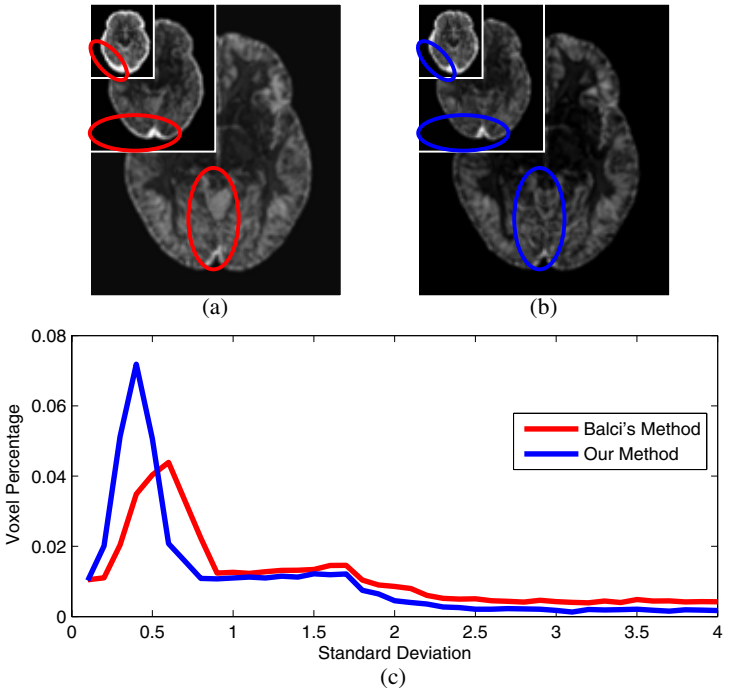


Fig. 2. Comparison of residual errors after groupwise registration via two different methods. (a) and (b): Our method (blue) produces lower errors than Balci's method (red), (c): The distributions of residual errors yielded by the two methods.

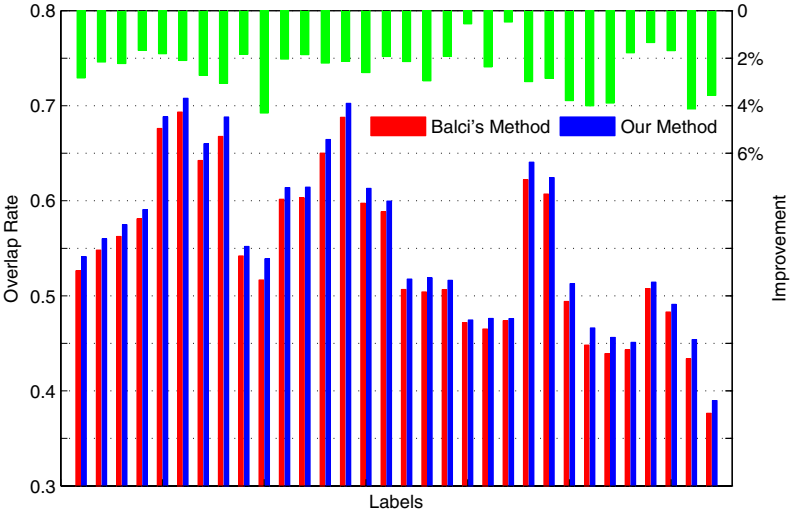


Fig. 3. Our method produces higher overlap rates than Balci's method in all 32 labels. The green bars represent the percentage of improvement by our method, relative to Balci's method.

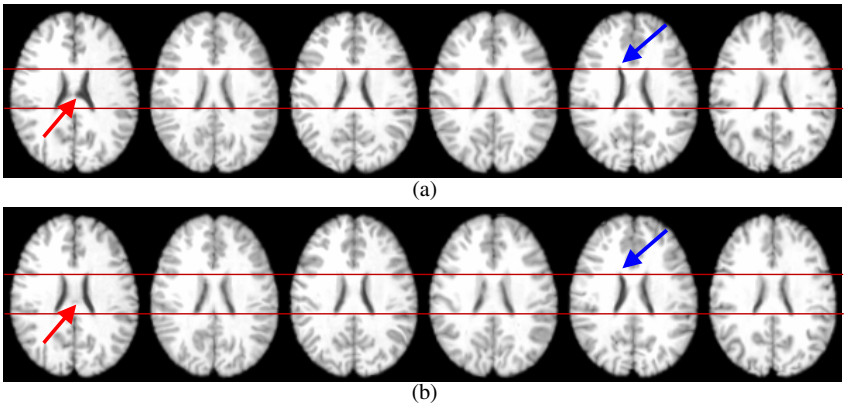


Fig. 4. Corresponding slices are selected from the images registered via Balci's method (a) and our method (b), respectively. Ventricles in row (b) are more similar to each other than those in row (a), especially for the locations indicated by arrows.

Furthermore, our method could better align anatomical structures in groupwise registration. For example, as shown in Fig. 4, we provide the same slices of 6 images selected from previously registered 16 images. The rows of (a) and (b) in Fig. 4 show the results produced via Balci's method and our method, respectively. As indicated by arrows, images registered by our method are much consistent than those yielded by Balci's method. In particular, ventricles in row (b) are more similar to each other than those in row (a).

4 Conclusion

In this paper, we have presented a novel groupwise registration method. Our method employs attribute vectors as anatomical signatures of voxels. We then estimate the PDF of each attribute at a certain voxel location from its neighborhood. Across different images, the dissimilarity of the PDF stack is further measured by the JS divergence. We can finally achieve groupwise registration by minimizing the overall JS divergence in a multi-resolution manner. Experimental results indicate that our method can yield better registration results than Balci's method, using the same implementation freely available in ITK.

Acknowledgments. This work is supported by NIH grants 1R01EB006733 and 1R01EB009634.

References

1. Zitová, B., Flusser, J.: Image Registration Methods: A Survey. Image and Vision Computing 21, 977–1000 (2003)
2. Crum, W.R., Hartkens, T., Hill, D.L.G.: Non-rigid Image Registration: Theory and Practice. The British Journal of Radiology 77, S140–S153 (2004)

3. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased Diffeomorphic Atlas Construction for Computational Anatomy. NeruoImage 23, S151–S160 (2004)
4. Miller, M.I., Younes, L.: Group Actions, Homeomorphisms, and Matching: A General Framework. International Journal of Computer Vision 41, 61–84 (2001)
5. Learned-Miller, E.G.: Data Driven Image Models through Continuous Joint Alignment. IEEE Trans. Pattern Analysis and Machine Intelligence 28, 236–250 (2006)
6. Zöllei, L., Learned-Miller, E., Grimson, E., Wells, W.: Efficient Population Registration of 3D Data. In: Liu, Y., Jiang, T.-Z., Zhang, C. (eds.) CVBIA 2005. LNCS, vol. 3765, pp. 291–301. Springer, Heidelberg (2005)
7. Balci, S.K., Golland, P., Wells, W.M.: Non-rigid Groupwise Registration using B-Spline Deformation Model. In: Workshop on Open-Source and Open-Data for 10th MICCAI, pp. 105–121 (2007)
8. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multi-modality Image Registration by Maximization of Mutual Information. IEEE Trans. Medical Imaging 16, 187–198 (1997)
9. Wells, W.M., Viola, P., Atsumi, H., Nakajima, S., Kikinis, R.: Multi-modal Volume Registration by Maximization of Mutual Information. Medical Image Analysis 1, 35–51 (1996)
10. Shen, D., Davatzikos, C.: HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration. IEEE Trans. Medical Imaging 21, 1421–1439 (2002)
11. Lin, J.: Divergence Measures Based on the Shannon Entropy. IEEE Trans. Information Theory 37, 145–151 (1991)
12. Wang, F., Vemuri, B.C., Rangarajan, A., Eisenschenk, S.J.: Simultaneous Nonrigid Registration of Multiple Point-Sets and Atlas Construction. IEEE Trans. Pattern Analysis and Machine Intelligence 30, 2011–2022 (2008)
13. Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H.: Introduction to the Non-rigid Image Registration Evaluation Project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 128–135. Springer, Heidelberg (2006)

Statistical Regularization of Deformation Fields for Atlas-Based Segmentation of Bone Scintigraphy Images

Karl Sjöstrand, Mattias Ohlsson, and Lars Edenbrandt

EXINI Diagnostics AB, Lund, Sweden
karl.sjostrand@exini.com

Abstract. The construction and application of statistical models of deformations based on non-rigid image registration methods have gained recent popularity. This paper presents the application of such a model to restricting a general-purpose registration algorithm to anatomically plausible solutions. Specifically, the Morphon registration method is used for atlas-based segmentation of bone scintigraphy images. From a training set of 734 images, a model of characteristic deformation fields is built and used for regularizing the registration of 113 test images. Results show that around 300 training images and 30 principal modes are sufficient for building a useful model. The segmentation succeeded in 106 of 113 test images.

1 Introduction

Accurate segmentation of bone scintigraphy images is a prerequisite for localizing and quantifying skeletal metastatic disease. Disease extent is an important prognostic indicator of survival longevity [1]. Automating the segmentation and the subsequent chain of analysis may both increase effectiveness and objectivity of the clinical investigation. In this paper, we describe and apply a modified version of the Morphon [2] algorithm for aligning a segmented atlas image to patient images. The principal modification pertains to the regularization of the vector field of deformations obtained during each iteration. To guide the registration towards anatomically plausible solutions, the deformation field is constrained with respect to a statistical deformation model [3] (SDM) built from examples. Previous reports on such models regarded small sets of examples and high-dimensional volumetric data [3,4]. Here, we focus on data of limited dimensionality and a relatively large training set to enable the investigation of the relations between model dimensionality, training set size, and generalization ability. The performance of the modified Morphon is measured from the segmentation of a test set of images.

The Morphon registration algorithm was devised by Knutsson and Andersson [2]. Petterson et al. [5] present an application of the Morphon to the segmentation of the pelvis from CT images. The term and concept of SDMs are due to

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 664–671, 2009.

Rueckert et al. [3]. SDMs have been applied directly to registration by optimization in the space of principal components [6,7,4], for the automatic construction of shape models [8,9] and for interpretation [10].

2 Method

This section presents a review of the Morphon method for image registration along with proposed modifications. Where applicable, we follow the nomenclature of Knutsson et al [2]. Image dimensionality is denoted by p_d while p is the number of image elements (pixels or voxels) and n is the number of images (subjects). The images studied in this paper are two-dimensional ($p_d = 2$), however, the discussed methodology applies to images of arbitrary dimensionality.

The Morphon method follows an iterative scheme where local image deformations are used to deform a source image with the aim of bringing it into successively closer correspondence with a target image. Instead of directly optimizing an image similarity measure, each Morphon iteration attempts to mitigate differences in location on a per-image element basis, subject to constraints on the irregularity of the resulting deformation field; a process which implicitly increases image similarity in successful cases. Each iteration consists of four steps; deformation of the source image according to the current accumulated deformation field, estimation of a new deformation field, deformation field regularization, and the addition of the regularized deformation field to the accumulated field. Each of these steps will be described briefly below.

Image Deformation. Each iteration starts with the deformation of the source image according to the current *accumulated field* (see below). This deformation is carried out using standard image warping techniques.

Deformation Field Estimation. A new deformation field is estimated from measurements of local phase differences between the deformed source image and the target image. Image phase is estimated using complex filters sensitive to intensity ridges, valleys and edges in a particular direction. A set of n_f filters (here, $n_f = 4$) are created to cover equally spaced directions on the unit circle. The complex filter response of each filter is separated into phase and magnitude components. The phase difference between the deformed source and target images at a particular element is proportional to the displacement of the element in the source image relative the target image. The magnitude of the response provides a measure of the certainty of the displacement. Using the filter directions as predictor variables and the estimated phase difference as the response, the deformation vector corresponding to an image element can determined by weighted least squares,

$$\arg\min_{\mathbf{v}} \sum_{i=1}^{n_f} c_i(\mathbf{n}_i^T\mathbf{v} - d_i)^2, \tag{1}$$

where $\mathbf{v}$ is the $p_d \times 1$ deformation vector to be estimated, $\mathbf{n}_i$ is the direction of the i^{th} filter, and d_i and c_i are the ditto phase response difference and magnitude respectively. Solving this equation for all image elements yields an initial deformation field estimate. We propose to augment Equation 1 to enable the registration of multi-band images such as RGB color images and multi-spectral and multi-modality images, where several images are acquired simultaneously and in correspondence. Bone scintigraphy is such an image type, as there are two corresponding images per patient, one anterior and one posterior. In such cases, the task is to determine a single deformation field from multiple images. Equation 1 can be augmented to include the additional information provided by several images in the following manner,

$$\arg\min_{\mathbf{v}} \sum_{j=1}^{n_b} \sum_{i=1}^{n_f} c_{i,j} (\mathbf{n}_{i,j}^T \mathbf{v} - d_{i,j})^2, \tag{2}$$

where n_b is the number of corresponding images (bands). This makes the size of the system $n_f n_b \times p_d$ instead of $n_f \times p_d$. The effect of this is that $\mathbf{v}$ can be determined based on more data, weighted by the certainty at each element, band and direction.

Deformation Field Regularization. The deformation field estimation neither models spatial dependencies among the deformation vectors nor enforces smoothness. Instead, this is incorporated in a subsequent regularization step on which much of this paper focuses. The Morphon method suggests a filtering approach known as *normalized averaging* resulting in elastic deformations [2,11]. Each component matrix of the deformation field is convolved with a Gaussian kernel. To increase robustness, the certainties are included which allow more certain deformations to have higher influence on the averaging at each element. Let the matrix $\mathbf{d}$ denote a component (e.g. the x-values) of the deformation field. Then, the regularized deformation field component is obtained by,

$$\mathbf{d}_{\text{reg}}^{\text{elastic}} = \frac{(\mathbf{c} \odot \mathbf{d}) * \mathbf{g}}{\mathbf{c} * \mathbf{g}}, \tag{3}$$

where $\mathbf{g}$ is a low-pass Gaussian filter kernel, $*$ is a convolution operator and $\odot$ is the Hadamard (element-wise) product operator. This procedure is carried out separately for all components.

During the first stages of the registration process, it may be beneficial to regularize the deformations further. We do this by projecting the deformation coordinates onto an affine basis in a weighted least-squares sense using the certainties as weights.

Deformation Field Accumulation. The obtained regularized deformation field is added to the total deformation field describing the transform from the original source image to the target. The certainties influence this accumulation

such that less certain deformations affect the the total field less than more certain counterparts. The accumulated field components $\mathbf{d}_a$ and the accumulated certainties $\mathbf{c}_a$ are updated in each iteration according to

$$\mathbf{d}_a = \frac{\mathbf{c}_a \odot \mathbf{d}_a + \mathbf{c} \odot (\mathbf{d}_a + \mathbf{d})}{\mathbf{c}_a + \mathbf{c}} \qquad \text{and} \qquad \mathbf{c}_a = \frac{\mathbf{c}_a^2 + \mathbf{c}^2}{\mathbf{c}_a + \mathbf{c}}. \tag{4}$$

We will now turn to a description of the proposed alternative to normalized averaging for displacement field regularization.

2.1 Building a Model of Characteristic Deformations

Using the described registration method we created an intensity-based atlas from ten bone scintigraphy images of normal subjects. The atlas was created as outlined in [12], where all subjects are registered to an arbitrary member of the normal group. The average deformation field is then calculated and used to transform all subjects to the common average anatomy. To decrease the bias from choosing a particular member as reference, the process is repeated with the estimated average anatomy and intensity as reference. The atlas was segmented into 31 anatomical regions by a medical expert. When registering an unknown bone image to this atlas, the atlas segmentation can be transformed using the inverse of the resulting deformation field, providing a segmentation of the unknown image.

We successfully registered 734 bone images to this atlas, resulting in equally many deformation fields. Affine transformation components were factored out, leaving a set of fields describing the non-rigid differences between each subject and the atlas. Each field was reshaped into a single row vector, e.g. as $[x_1 \dots x_p \; y_1 \dots y_p]$ and put in a mean centered data matrix $\mathbf{X}$ of size $n \times p_d p$. Performing a principal component analysis (PCA) on these data amounts to finding and orthogonal matrix $\mathbf{L}$ and a diagonal matrix $\mathbf{D}$ such that $n^{-1}\mathbf{X}^T\mathbf{X}\mathbf{L} = \mathbf{L}\mathbf{D}$. However, as $p \gg n$ for most SDMs, the computation of the $p_d p \times p_d p$ covariance matrix $n^{-1}\mathbf{X}^T\mathbf{X}$ becomes difficult. The principal components and their variances can instead be obtained by

$$\mathbf{L} = \frac{1}{\sqrt{n}}\mathbf{X}^T\hat{\mathbf{L}}\hat{\mathbf{D}}^{-\frac{1}{2}}, \qquad \mathbf{D} = \hat{\mathbf{D}}, \tag{5}$$

where $\hat{\mathbf{L}}$ and $\hat{\mathbf{D}}$ are the eigenvectors and eigenvalues respectively of the smaller $n \times n$ matrix $n^{-1}\mathbf{X}\mathbf{X}^T$.

The deformation fields are defined on a rectangular grid of which only part is occupied by the object of interest, the skeleton. The decomposition of the deformation field data can be done as suggested above, but much of the model dimensionality will be spent on describing deformations which occur outside the skeleton. This can be alleviated by performing an importance-weighted PCA [13]. We used the mean certainty map calculated from all 734 training images as weights and put them (repeated p_d times) in a diagonal matrix $\mathbf{W}$. The decomposition can then be performed as above using the weighted data set $\tilde{\mathbf{X}} = \mathbf{X}\mathbf{W}$.

2.2 Model-Based Regularization of Deformation Fields

Regularization using normalized averaging is general in the sense that it allows any transformation with sufficient smoothness. Many such transformations lead to anatomically implausible solutions. Regularization can instead be provided by the model of characteristic transformations from Equation 5. The regularization consists of a projection of the obtained deformation field $\mathbf{d}$ (here represented by a length $p_d p$ column vector) onto a rank k subspace by

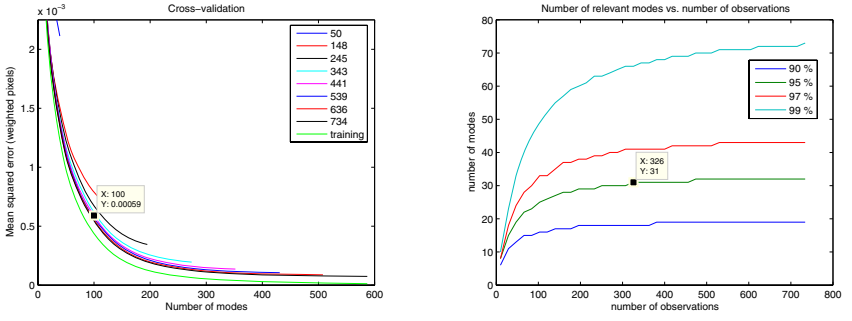
$$\mathbf{d}_{\text{reg}}^{\text{SDM}} = \mathbf{W}^{-1}\mathbf{L}_k\mathbf{L}_k^T\mathbf{W}(\mathbf{d} - \mathbf{d}_{\text{affine}}) + \mathbf{d}_{\text{affine}}, \tag{6}$$

where $\mathbf{L}_k$ consists of the k columns of $\mathbf{L}$ corresponding to the k largest eigenvalues and $\mathbf{d}_{\text{affine}}$ is the affine component of $\mathbf{d}$.

3 Results

To determine whether the given number of training images is sufficient for building a model that is general enough for segmenting unseen images, we performed two studies. The first is a five-fold cross-validation study measuring the discrepancy between the "true" deformation fields as given by standard Morphon segmentation of the 734 training set cases and the model approximation of the corresponding fields. The error measure was $s^{-1}\sum_{i=1}^{s}\|\mathbf{d}_i - \mathbf{d}_{i(\text{reg})}^{\text{SDM}}\|^2$ where $s = \lfloor 734/5 \rfloor$ is the number of images in the test fold. This measure was evaluated for an increasing number of modes and for eight data set sizes; n = 50, 148, 245, 343, 441, 539, 636 and 794. We also evaluated the measure without cross-validation, i.e. directly on the training data. Figure 1(a) shows the resulting eight cross-validation error curves and the training error curve. From a computational viewpoint, cross-validation is impractical for high-dimensional data sets; the study presented here took days to compute. A less demanding alternative is to, as is customary, assume that the relevant modes explain e.g. 95 % of the total variation. Focusing on the eigenvalues of submatrices of the "small" covariance matrix, we plot the number of modes necessary to explain a certain fraction of the total variance given a data set size. Figure 1(b) shows this plot for fractions 90 %, 95 %, 97 % and 99 %. The model size saturate at around 30 modes and 300 training images. The cross-validation study also shows scant improvement for models built from more that 300 images, but seems to suggest a larger model (around 100 modes). Parallel analysis [14] performed on repetitions of permuted variables and all 734 images suggested 27 modes. Based on this information and favoring a more compact model, we chose a model of 30 modes built from all 734 images for the subsequent analysis.

Both the standard Morphon registration and the SDM-regularized registration were initiated with an affine alignment to account for gross differences. The SDM registration was then regularized using an increasing number of modes; 1, 2, 5, 10, and finally 30. The algorithm was run for a fixed number of iterations (typically 5) for each of these regularization options. This proved sufficient for convergence. The SDM algorithm was run on a test set of 113 unique images/patients. Visual



(a) 5-fold cross-validation

(b) Training set size versus suitable model size

Fig. 1. Curves for determining model dimensionality and the smallest training set size for which the addition of more training examples does not lead to a more general model. Both graphs indicate that models built from around 300 examples are sufficiently general, but cross-validation suggests models of higher dimensionality than the analysis of the principal component variances.

inspection yielded 7 failures, 6 with minor registration errors and 1 with more pronounced errors. The standard algorithm, which already had proven its value by succeeding in 734 of 795 training images, failed in 10 cases from the test set, 8 and 2 with minor and major errors respectively.

4 Discussion

Figure 2 shows the atlas with the manually defined segmentation superimposed along with examples of images where SDM regularization managed to guide the registration to a better segmentation than normalized averaging regularization. Example 1 shows a patient with a leg prosthesis which is not detected on scintigraphy images. SDM regularization leads to a more plausible segmentation in this region compared to standard regularization. The right arm is also better delineated with SDM regularization in this example. Example 2 shows an example where the skull segmentation was placed relatively far from from the actual skull after affine alignment. SDM regularization manages to preserve the skull shape and guide the segmentation into place while standard regularization leads to a smearing of the skull segmentation such that only part of the delineation is correct. The choice of incorporating a maximum of 30 modes for regularization was sufficient for obtaining accurate registrations on the test images. Raising this number to 100 as suggested by Figure 1(a) neither improved nor over-fitted the adjusted model to a significant extent.

The segmentation failure rates for SDM versus standard regularization suggests that performance is comparable between the methods. Often, the failures occurred for different patients, suggesting SDM regularization as a complement — rather than as a replacement — to normalized averaging.

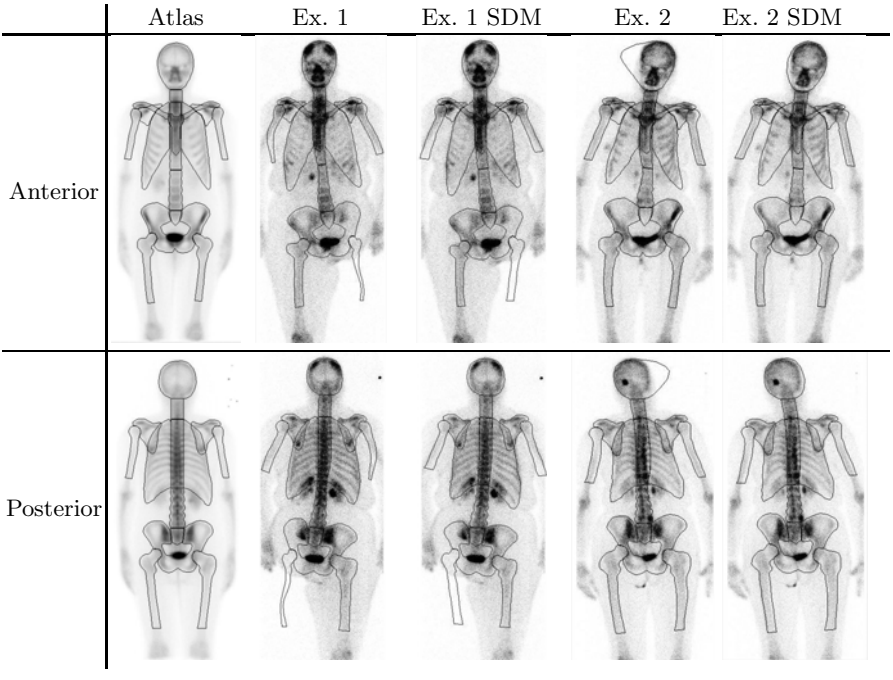


Fig. 2. Anterior and posterior bone scintigraphy images with the segmentation superimposed. The first column shows the atlas with the manually drawn segmentation. Columns 2 and 3 show example results from registering the atlas to a patient image using normalized averaging and SDM regularization respectively. Columns 4 and 5 show another example.

The Morphon registration method is based mainly on convolutions. Optimizing the algorithm for speed therefore amounts to creating an efficient convolution procedure. The registration of one patient containing one anterior and one posterior image took roughly 5-10 seconds in our Java implementation, for both normalized averaging and SDM regularization. The computational complexity for these regularization options is $\mathcal{O}(lp)$ and $\mathcal{O}(kp)$ respectively, where l is the size of the (1-D) regularization kernel and k is the number of modes.

The choice of decomposing the deformation data by an importance-weighted PCA had a large impact on model compactness. Performing a standard PCA lead to a model which required 73 modes to cover 95 % of the variation while parallel analysis suggested 54 modes. The graphical method from Figure 1(b) suggested around 70 modes calculated from at least 500 training images, however, the curves did not flatten out as evidently as for the weighted decomposition. In general, our impression is that the permeating use of the certainty matrix in the Morphon algorithm makes a difference and sets it apart from other optical flow-type algorithms for registration.

In studies to come, we wish to investigate the possibility of using the described method for registering volumetric image data. Previous reports on SDMs have presented preliminary results on this [7,4]. However, in light of the results

presented here, the outlook of gathering enough data for building a sufficiently general model seems bleak for such complex data sets.

References

1. Miller, P.D., Eardley, I., Kirby, R.S.: Prostate specific antigen and bone scan correlation in the staging and monitoring of patients with prostatic cancer. British Journal of Urology 70, 295–298 (1992)
2. Knutsson, H., Andersson, M.: Morphons: Paint on priors and elastic canvas for segmentation and registration. In: Scandinavian Conference on Image Analysis, SCIA (2005)
3. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3D statistical deformation models using non-rigid registration. IEEE Transactions on Medical Imaging 22, 77–84 (2003)
4. Thompson, S., Penney, G., Buie, D., Dasgupta, P., Hawkes, D.: Use of a CT statistical deformation model for multi-modal pelvic bone segmentation. In: SPIE Medical Imaging 2008: Image Processing, vol. 6914, pp. 9141–9141. SPIE (2008)
5. Pettersson, J., Knutsson, H., Borga, M.: Automatic hip bone segmentation using non-rigid registration. In: Proceedings of the IEEE International Conference on Pattern Recognition, Hong Kong, China (2006)
6. Loeckx, D., Maes, F., Suetens, P.: Temporal subtraction of thorax CR images using a statistical deformation model. IEEE Transactions on Medical Imaging 22, 1490–1504 (2003)
7. Wouters, J., D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P.: Non-rigid brain image registration using a statistical deformation model, vol. 6144., 614411. SPIE (2006)
8. Frangi, A.F., Rueckert, D., Schnabel, J., Niessen, W.J.: Automatic 3D ASM construction via atlas-based landmarking and volumetric elastic registration. In: Insana, M.F., Leahy, R.M. (eds.) IPMI 2001. LNCS, vol. 2082, pp. 78–91. Springer, Heidelberg (2001)
9. Heitz, G., Rohlfing, T., Maurer Jr., C.R.: Statistical shape model generation using nonrigid deformation of a template mesh. In: Fitzpatrick, J.M., Reinhardt, J.M. (eds.) Proceedings of the SPIE, Medical Imaging 2005: Image Processing, vol. 5747, pp. 1411–1421. SPIE (2005)
10. Ólafsdóttir, H., Hansen, M.S., Sjöstrand, K., Darvann, T.A., Hermann, N.V., Oubel, E., Ersbøll, B.K., Larsen, R., Frangi, A.F., Larsen, P., Perlyn, C.A., Morriss-Kay, G.M., Kreiborg, S.: Sparse statistical deformation model for the analysis of craniofacial malformations in the crouzon mouse. In: Ersbøll, B.K., Pedersen, K.S. (eds.) SCIA 2007. LNCS, vol. 4522, pp. 112–121. Springer, Heidelberg (2007)
11. Bro-Nielsen, M.: Medical Image Registration and Surgery Simulation. PhD thesis, Department of Mathematical Modelling, Technical University of Denmark (1996)
12. Guimond, A., Meunier, J.: Average brain models: A convergence study. Computer Vision and Image Understanding 77(77), 192–210 (2000)
13. Jolliffe, I.T.: Principal Component Analysis, 2nd edn. Springer, Heidelberg (2002)
14. Horn, J.L.: A rationale and a test for the number of factors in factor analysis. Psychometrika 30, 179–185 (1965)

Graphical Models and Deformable Diffeomorphic Population Registration Using Global and Local Metrics

Aristeidis Sotiras[1,2], Nikos Komodakis[3], Ben Glocker[1,4], Jean-François Deux[5], and Nikos Paragios[1,2]

[1] Laboratoire des Mathématiques Appliquées aux Systèmes (MAS)
Ecole Centrale de Paris, France
aristeidis.sotiras@ecp.fr
[2] Equipe GALEN, INRIA Saclay - Île de France, Orsay, France
[3] Computer Science Department, University of Crete, Greece
[4] Computer Aided Medical Procedures (CAMP), Technische Universität München
Munich, Germany
[5] Service d'Imagerie Médicale, Centre Hospitalier Universitaire Henri Mondor, Créteil, France

Abstract. In this paper we propose a novel framework to unite a population to an optimal (unknown) pose through their mutual deformation. The registration criterion comprises three terms, the first imposes compactness on appearance of the registered population at the pixel level, the second tries to minimize the individual distances between all possible pairs of images, while the last is a regularization one imposing smoothness on the deformation fields. The problem is reformulated as a graphical model that consists of hidden (deformation fields) and observed variables (intensities). A novel deformation grid-based scheme is proposed that guarantees the diffeomorphism of the deformation and is computationally favorably compared to standard deformation methods. Towards addressing important deformations we propose a compositional approach where the deformations are recovered through the sub-optimal solutions of successive discrete MRFs by using efficient linear programming. Promising experimental results using real 2D data demonstrate the potentials of our approach.

1 Introduction

Population registration is defined as the identification of a homology between more than two images. Its importance is evident in problems like statistical modeling of variations and atlas construction. To solve the fore-mentioned problems, often a reference frame is chosen and all population members are mapped to this pose using pair-wise registration algorithms. The explicit selection of the reference image bias inherently the registration towards the chosen reference frame [1] and influence inherently its performance. Such a behavior is the opposite to the one expected towards appropriate representation of the population. Last but not least, these methods are not applicable when aiming statistical deformations population modeling using data coming from different modalities.

Methods that try to overcome the above-stated limitations can be subdivided into two classes. The first class of methods initially focuses on the appropriate selection of the reference [2] or constructs a reference template through the use of the population statistics [3,1]. Conventional registration methods based in pair-wise criteria are

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 672–679, 2009.

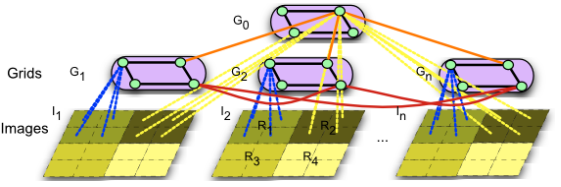


Fig. 1. The node and the edge system of the constructed graph. With blue color the relationship between the grid nodes and the images is depicted (deformation model). The black edges represent the smoothness terms while the red ones encode the local dissimilarity measure. The global relationship between all the nodes at respective places in the grids is shown by the yellow edges. (For clarity a fraction of the edges is shown.)

then considered towards the reference. The main limitation of these methods lies in the use of the template. The second class are template-free group-wise registration methods using either local pairwise relations or global population measurements. In [4], all possible pair-wise registrations were considered and a mean model was created by composing the deformations for each member into a mean deformation. In [5], the sum of univariate entropies along pixel stacks is introduced to address the problem of group-wise registration using an affine deformation model and is further extended by [6] to include FFD. Last but not least, in [7] local pair-wise relations were considered to deform mutually a population of images towards providing an atlas-based segmentation.

Template-driven methods introduce bias to the process through the selection of the reference, and treat individually examples of the population. On the other hand, template-free population-registration methods suffer from the lack of modularity with respect to the registration criterion and the deformation model, are sensitive to the initial conditions while being computationally inefficient. The scope of the objective function is limited to pairwise relations and computational approximations [5] are used in order to meet the high computational and memory demands.

In this paper, we propose a graphical model approach to population registration [Fig.1]. The latent variables of the model are$n$-deformations (Hermite-based polynomials) of the population examples and the optimal reference pose. The pose variables are connected with the observations and the corresponding deformation variables towards measuring the statistical compactness of the registration result at the pixel level. The registration variables are inter-connected and aim to decrease the cost of pair-wise comparisons between individual examples. Last, but not least the registration variables within an image are connected so as to impose smoothness. The resulting paradigm can easily encode different deformation interpolation methods, local similarity metrics and global statistical measurements while being computational efficient [when compared with the state of the art methods]. This graphical model is expressed in the form of a MRF. Towards validating the approach, we consider population registration of calf muscle MRI images.

2 Global and Local Population Registration

Let us consider n images $\{I_1, ..., I_n\}$, where each image is described by intensity values $I_i(\mathbf{x}_i)$ for different image domains $\Omega_i, \mathbf{x}_i \in \Omega_i$. The aim of the mutual population

deformation is to determine a set of transformations $\mathbf{T} = \{T_i : \mathbf{x}_R = T_i(\mathbf{x}_i), i = \{1, ..., n\}\}$ which maps mutually corresponding points from the n-image spaces to the same point of a reference frame Ω_R. In our case, we assume the reference pose to simply correspond to the geometry and not an image template.

Deformation Model. Let us consider a grid-based deformation model that can encode different interpolation methods in a way that the transformation is one-to-one and invertible. The deformation of an object is achieved by manipulating an underlying mesh of control points. We superimpose a deformation grid $G_i : [1, K] \times [1, L]$ onto each one of the images I_i and let us also consider a grid $G_0 : [1, K] \times [1, L]$ in the reference pose. The central idea of our approach is to deform the grids simultaneously (with a given displacement vector $\mathbf{d}_{\mathbf{p}_i^k}$ for each control point k belonging to the grid G_i) such that meaningful correspondences between the population examples are obtained and their mapping to the reference pose creates a statistically compact variable. In this case, the transformation of an image pixel $\mathbf{x}_i = (x_i, y_i) \in \Omega_i$ can be written as $T_i(\mathbf{x}_i) = \mathbf{x}_i + D_i(\mathbf{x}_i)$ where $D_i(\mathbf{x}_i) = \sum_{\mathbf{p}_i^k \in G_i} \eta(|\mathbf{x}_i - \mathbf{p}_i^k|)\mathbf{d}_{\mathbf{p}_i^k}$ and $\eta(\cdot)$ is a weighting function that measures the contribution of the control point $\mathbf{p}_i^k$ to the displacement field D_i.

Population-wise Global Comparisons. The first term of the objective criterion to be minimized is the global statistical compactness one. We consider the intensity values of the deforming images at corresponding coordinate locations as a distribution of a random variable $\pi(\mathbf{i}(\mathbf{x}))$, where $\mathbf{i}(\mathbf{x}) = \{I_1(T_i^{-1}(\mathbf{x})), \cdots, I_n(T_n^{-1}(\mathbf{x}))\}$. In statistics, one can associate a random variable to a measure of compactness with respect to this density. Examples can refer to standard deviation, higher order moments, Shannon entropy, etc. It should be expected that as the images are aligned the compactness of the probability distribution should increase. We introduce the following global measurement towards population registration

$$E_g(T_1, \cdots, T_n) = \iint_{\Omega_R} \gamma(\pi(\mathbf{i}(\mathbf{x})))d\mathbf{x} \tag{1}$$

with γ being a monotonic function inversely proportional to the compactness of the intensity distribution at $\mathbf{x}$ once all population examples have been mapped to the reference pose. Such an objective function introduces the inverse transformation, that is challenging from theoretical and practical point of view when referring to deformable deformation. An alternative criterion that can be considered is using the forward transformations and measure the similarity of the images on the intersection of the deformed images, or

$$E_g(\mathbf{T}) = \int \cdots \int_{\Omega_i \cup \cdots \cup \Omega_n} \phi(T_1(\mathbf{x}_1), \cdots, T_n(\mathbf{x}_n))\gamma(\lambda(\mathbf{x}_1, \cdots \mathbf{x}_n))d\mathbf{x}_1 \cdots d\mathbf{x}_n \tag{2}$$

where $\lambda(\mathbf{x}_1, \cdots \mathbf{x}_n) = \pi(I_1(T_1(\mathbf{x}_1)), \cdots, I_n(T_n(\mathbf{x}_n)))$ and ϕ is a Dirac-driven function whose role is to define which pixels correspond to the same position at the reference pose defined as follows: $\prod_{(i,j)\in[1,n]\times[1,n]} \delta_\alpha(|\mathbf{x}_i - \mathbf{x}_j|)$.

Population-wise Local Comparisons. It may be the case that a distribution exhibits good compactness characteristics globally but certain members of the population can always be placed in the tale of the distribution. To avoid such cases, local pair-wise

comparison between the members of the population are going to be considered. Let $\rho_{ij}(\cdot)$ be a similarity measurement used to compare the visual information for the images i and j. Then, if (without loss of generality) we consider for example pixel-based measurements, the pair of forward deformations T_i, T_j, should minimize the distance in the intersection of the deformed images:

$$E_l(T_i, T_j) = \iint_{\Omega_i \cup \Omega_j} \phi(|T_i(\mathbf{x}_i) - T_j(\mathbf{x}_j)|)\rho_{ij}(I_i(T_i(\mathbf{x}_i)), I_j(T_j(\mathbf{x}_j)))d\mathbf{x}_i d\mathbf{x}_j \quad (3)$$

In simple words, this quantity evaluates the pertinence of the correspondences between the two images using both definition domains Ω_i, Ω_j where only the pixels for which correspondences between the two images have been found are considered. The criterion can be extended to deal with the case of n-images by simply considering all possible pairs of images.

Smoothness Constraints. Medical images capture properties of spatially continuous anatomical structures, therefore it is natural to assume that the deformation applied to them should be locally smooth. Opposite to the former cases, this constraint should applied to each grid separately. This constraint can be defined on the grid as

$$E_s(T_1, \cdots, T_n) = \sum_{i=1}^{n} \iint_{\Omega_i} \psi(\nabla_{T_i}(\mathbf{x_i}))d\mathbf{x}_i \quad (4)$$

where ψ is a convex function imposing smoothness.

The optimal parameters of the deformation should be determined through the minimization of an objective function being composed of the above terms. Gradient descent method is the most common approach, but is unable to guarantee the recovery of the global minimum, is computational inefficient, and far from being modular. Graphical models and the off-the-shelf discrete optimization methods being associated to them can address the above mentioned constraints.

3 Graphical Model towards Population Registration

In order to able to use discrete optimization schemes the deformation space should be quantized. Let $\Theta = \{\mathbf{d}^1, .., \mathbf{d}^q\}$ be a quantized version of the deformation field, then a discrete set of labels $L = \{l^1, ..., l^q\}$ can be corresponded to it. A label assignment $l^{\xi}_{\mathbf{p}}$, where $\xi \in \{1, \cdots, q\}$, to a grid node $\mathbf{p}$ is associated with displacing the node by the corresponding vector $\mathbf{d}^{l^{\xi}_{\mathbf{P}}}$. If a label is assigned to every node we get a discrete labeling $\mathbf{l}$. The displacement field associated with a certain labeling $\mathbf{l}$ becomes $D(\mathbf{x}) = \sum_{\mathbf{p} \in G} \eta(|\mathbf{x} - \mathbf{p}|)\mathbf{d}^{l_{\mathbf{P}}}$. We have considered the Hermite splines. In this case $D = \sum_{l=0}^{1}\sum_{m=0}^{1} H_l(u)H_m(v)\mathbf{d}_{i+l,j+m}$. $i = \lfloor x/\delta_x \rfloor, j = \lfloor y/\delta_y \rfloor, u = x/\delta_x - \lfloor x/\delta_x \rfloor$ and $v = y/\delta_y - \lfloor y/\delta_y \rfloor$. H_l represents the lth basis function of the Hermite spline and $\delta_x = \frac{M}{K-1}, \delta_y = \frac{N}{L-1}$ denotes the control point spacing. Hermite splines involve less computations than cubic-B splines while exhibiting the same desired properties.

By applying this quantization of the deformation space one would like to reformulate the problem as a discrete multi-labeling problem. A common model for representing

such problems are Graphical Models and MRFs. In the context of population registration, the graphical model will involve three terms, one singleton that measures the compactness and two pair-wise, one that account for smoothness at each deformation field and one that enforces pair-wise correspondences.

$$E_{GM}(G_0, T_1 \circ G_1, \cdots, T_n \circ G_n,) = \alpha \sum_{i=0}^{n} \sum_{\mathbf{p} \in G_0} V_{\mathbf{p}}(l_{\mathbf{p}}) + \\ \beta_{intra} \sum_{i=0}^{n} \sum_{\mathbf{p} \in G_i} \sum_{\mathbf{q} \in (N(\mathbf{p}) \cap G_i)} V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}}) + \beta_{inter} \sum_{i=0}^{n} \sum_{\mathbf{p} \in G_i} \sum_{\mathbf{q} \in (N(\mathbf{p}) \setminus G_i)} V_{\mathbf{pq}}(l_{\mathbf{p}}, l_{\mathbf{q}}) \tag{5}$$

where $V_{\mathbf{p}}(\cdot)$ are the unary potentials, $V_{\mathbf{pq}}(\cdot, \cdot)$ are the pair-wise potentials and N represents the neighborhood system of the nodes [Fig.1]. α, β_{inter} and β_{intra} are weighting constants. The main challenge of discrete optimization methods is the quantization of the search space since it seeks for a compromise between computational complexity and the ability to capture a good minimum. This can be achieved through a compositional approach, where the final solution is obtained through successive optimization problems with respect to the deformation increment towards minimizing the objective function [8]. Thus, by keeping the set of the labels in a reasonable size it becomes possible to approximate the optimal solution in an efficient way.

3.1 Mapping of the Objective Function to the Graphical Model

Mapping global, local and smoothness costs to the graphical model consists of converting them to singleton and pair-wise terms. The most challenging case is the global cost due to the fact that in order to be properly determined it requires higher order cliques. The mapping of the other two terms is straightforward.

Singleton Term. The adoption of higher order cliques is possible within MRFs, however their use decreases significantly their computational efficiency. We consider an approximation of the global cost that consists of assuming that for a given node $\mathbf{p}$ of a given deformation field/image i, the rest of the images do not move within the current iteration. This assumption is considered for all nodes, and for all deformation fields within a given iteration and therefore is not restrictive and quite common in minimizing graphical models through expansion moves. Then, the cost of a deformation will depend only on the label of this node, or,

$$V^t_{\mathbf{p_i^k}}(l_{\mathbf{p_i^k}}) \approx \int \cdots \int_{\Omega_1 \cup \cdots \cup \Omega_n} \eta_s^{-1}(\mathbf{x}_i, \mathbf{p}_i^k) \phi(T_1^{t-1}(\mathbf{x}_1), \cdots, T_i^t(\mathbf{x}_i), \cdots T_n^{t-1}(\mathbf{x}_n)) \\ \gamma(\lambda(I_1(T_1^{t-1}(\mathbf{x}_1)), \cdots, I_i(T_i^t(\mathbf{x}_i)))) d\mathbf{x}_1 \cdots d\mathbf{x}_n \tag{6}$$

where $\eta_s^{-1}(\mathbf{x}_i, \mathbf{p}_i^k) = \eta_s^{-1}(|\mathbf{x}_i - \mathbf{p}_i|) = \frac{\eta(|\mathbf{x}_i - \mathbf{p}_i|)}{\int_{\Omega_i} \eta(|\mathbf{y}_i - \mathbf{p}_i|) d\mathbf{y}_i}$. We have considered a congealing-like global cost that aims at minimizing the entropy of the pixel distributions upon registration. This term corresponds to the G_0 graphical model variables.

Pair-Wise Terms. Two different cases have to be discerned, one that accounts for pair-wise registrations between all image pairs and one that imposes smoothness on the deformation fields. The adaptation of the local registration costs involves connections

between the nodes $\mathbf{p}_i^k, \mathbf{q}_j^k$, that are placed in respective places k in grids that belong to two different images i and j. The inter pair-wise potential are defined as

$$V_{\mathbf{p}_i^k \mathbf{q}_j^k}(l_{\mathbf{p}_i^k}, l_{\mathbf{q}_j^k}) \approx \int_{\Omega_i \cup \Omega_j} \eta_p^{-1}(\mathbf{x}_i, \mathbf{p}_i^k, \mathbf{x}_j \mathbf{q}_j^k) \phi(|T_i(\mathbf{x}_i) - T_j(\mathbf{x}_j)|) \cdot \rho(I_i(T_i(\mathbf{x}_i)), I_j(T_j(\mathbf{x}_j))) d\mathbf{x}_i d\mathbf{x}_j \tag{7}$$

where η^{-1} are inverse projection functions that depend on the distances between the pixel and the different deformation grids and are defined as: $\eta_p^{-1}(\mathbf{x}_i, \mathbf{p}_i^k, \mathbf{x}_j \mathbf{q}_j^k) = \frac{\eta(|\mathbf{x}_i - \mathbf{p}_i|)\eta(|\mathbf{x}_j - \mathbf{q}_j|)}{\int_{\Omega_i \cup \Omega_j} \delta(T_i(\mathbf{y}_i), T_j(\mathbf{y}_j))\eta(|\mathbf{y}_i - \mathbf{p}_i|\eta(|\mathbf{y}_j - \mathbf{q}_j|)) d\mathbf{y}_i d\mathbf{y}_j}$. The image metric used in the context of pair-wise image comparisons of our approach was the sum of absolute differences.

Last, but not least imposing smoothness on the deformation fields can be done by defining a distance function computing the magnitude of vector differences [8]

$$V_{\mathbf{p_i q_i}}(l^{\xi}_{\mathbf{p}_i}, l^{\nu}_{\mathbf{q}_i}) = |\mathbf{d}^{l^{\xi}_{\mathbf{p}_i}} - \mathbf{d}^{l^{\nu}_{\mathbf{q}_i}}|.$$

To minimize the successive MRFs, that is to assign a label l to all the nodes $\mathbf{p}$ of the constructed graph, an efficient linear programming method is used [9]. The last constraint to be addressed refers to the diffeomorphic property of the proposed population registration framework. This can easily introduced by imposing hard constraints to the allowed deformations [10]. Following [11], the bound for the maximum displacement towards guaranteeing diffeomorphic deformations, in the case of the cubic Hermite spline, is proven to be 0.25 times the grid spacing.

4 Experimental Validation

To evaluate the performance of the method, the population registration of 2D MR human skeletal muscle calf images has been considered. The images were acquired with a 1.5T Siemens scanner, with parameters TR=711, TE=11. Each volume consists of 90 slices of 4mm thickness with voxel spacing of $0.7812 \times 0.7812 \times 4$ mm. From the

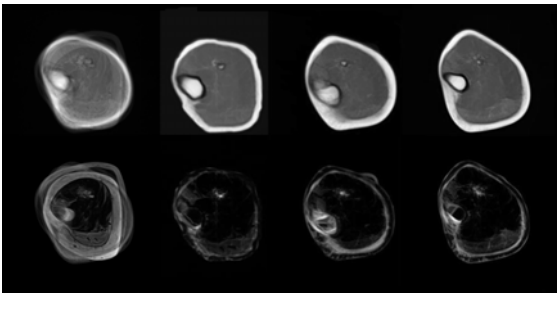


Fig. 2. Results obtained for the muscle image data set (mean and variance image). From left to right, the initial images, the result of the group-wise registration, the result of two pair-wise registrations.

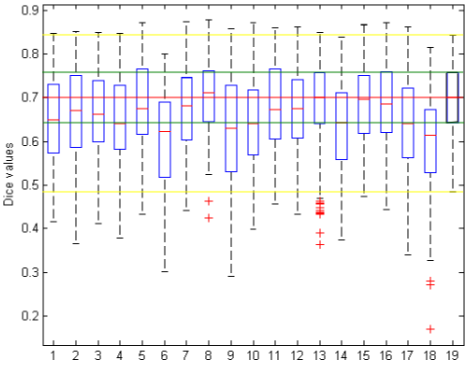


Fig. 3. Comparison between group-wise and pair-wise registration. The Dice coefficients obtained through pair-wise registration with respect to all plausible individual template choices are compared with the population registration result.

original volumes slices that correspond to respective positions were selected to for the data set. Segmentations for the data were provided by an expert. The parameters of the method, $\alpha, \beta_{inter}, \beta_{intra}$ were set to 10, 1 and 0.1, respectively. We have used a multi-scale implementation with 3 levels, an initial grid resolution of 8×8, and a final one of 32×32. A number of $2 \times 4 + 1$ labels were used per iteration cycle, sampled along the principal horizontal and vertical directions.

The qualitative results of the group-wise registration of the muscle data are presented in [Fig.2]. Comparing visually the mean and the variance image of the population before and after the group-wise registration the success of the registration process can be assessed qualitatively. The mean image is far more sharp than the one before the registration process, while the variance image emphasizes the decrease of the intensity differences along the registered data.

To further appraise the performance of the proposed method, it was compared to a state of the art pair-wise registration method [8]. Similar parameters and deformation grids were used for both methods with the difference that for the group-wise registration scheme, Hermite weighting functions were used instead of cubic B-splines. The performance of the pairwise registration was exhaustively evaluated as all possible images were used as targets. The distributions of the Dice values for each image target are reported in [Fig.3], where a boxplot is given for every image target. The results for the pair-wise registration are given from column 1 to 18, while the last column corresponds to the results obtained by the proposed group-wise registration framework. By simple observation, it can be concluded that the group-wise registration outperforms the pair-wise method for the majority of the cases.

The results depicted in the graph suggest that considering the population as a whole and registering subjects jointly brings the population into better alignment than matching each subject to a target image. This is implied by the decrease of the dispersion of the Dice values that is observed in the group-wise case. The results presented in the figure [Fig. 3] point out the intrinsic drawbacks of the pair-wise registration process whose performance is greatly influenced by the choice of the target image.

A Matlab implementation of our approach takes approximately 30 min on an Apple Mac with 4GB memory and 2.5GHZ Processor, for a population registration of 20 examples (256×256) and a final resolution grid of 32×32 per image. However, since our graph is similar to the one in [8] and the same optimization technique is used, a C++ implementation should decrease the running time to a couple of minutes.

5 Discussion

In this paper we have proposed a novel approach to unbiased diffeomorphic deformable population registration using graphical models and discrete optimization. Our approach is gradient free, modular in terms of the image and smoothness components and can encode global population criteria and pair-wise comparisons.

The extension of the method to deal with 3D data is natural and straightforward future direction. Furthermore, the use of higher-order MRFs towards proper approximation of the global costs will improve the performance of the method in terms of ability to capture the global optimum. Last, but not least the ability to construct an unbiased statistical anatomical atlas using the proposed concept could be a useful tool in a number of applications in medical imaging.

References

1. Bhatia, K., Hajnal, J., Puri, B., Edwards, A., Rueckert, D.: Consistent groupwise non-rigid registration for atlas construction. In: ISBI (2004)
2. Park, H., Bland, P.H., Hero III, A.O., Meyer, C.R.: Least biased target selection in probabilistic atlas construction. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 419–426. Springer, Heidelberg (2005)
3. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage (2004)
4. Seghers, D., D'Agostino, E., Maes, F., Vandermeulen, D., Suetens, P.: Construction of a brain template from MR images using state-of-the-art registration and segmentation techniques. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 696–703. Springer, Heidelberg (2004)
5. Zollei, L., Learned-Miller, E., Grimson, E., Wells, W.: Efficient population registration of 3d data. In: ICCV (2005)
6. Balci, S., Golland, P., Shenton, M., Wells, W.: Free-form b-spline deformation model for groupwise registration. In: MICCAI (2007)
7. Sotiras, A., Komodakis, N., Langs, G., Paragios, N.: Atlas-based deformable mutual population segmentation. In: ISBI (2009)
8. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through mrfs and efficient linear programming. In: MIA (2008)
9. Komodakis, N., Tziritas, G., Paragios, N.: Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies. In: CVIU (2008)
10. Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J.V., Hammers, A.: Diffeomorphic registration using B-splines. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 702–709. Springer, Heidelberg (2006)
11. Choi, Y., Lee, S.: Injectivity conditions of 2d and 3d uniform cubic b-spline functions. Graphical Models (2000)

Efficient Large Deformation Registration via Geodesics on a Learned Manifold of Images*

Jihun Hamm, Christos Davatzikos, and Ragini Verma

Department of Radiology, University of Pennsylvania, USA
{hammj,christos.davatzikos,ragini.verma}@uphs.upenn.edu

Abstract. Geodesic registration methods have been used to solve the large deformation registration problems, which are hard to solve with conventional registration methods. However, analytically defined geodesics may not coincide with anatomically optimal paths of registration. In this paper we propose a novel and efficient method for large deformation registration by learning the underlying structure of the data using a manifold learning technique. In this method a large deformation between two images is decomposed into a series of small deformations along the shortest path on the graph that approximates the metric structure of data. Furthermore, the graph representation allows us to estimate the optimal group template by minimizing geodesic distances. We demonstrate the advantages of the proposed method with synthetic 2D images and real 3D mice brain volumes.

1 Introduction

The problem of registering two images is particularly challenging in the presence of large deformations between the images, as in the case of growth in pediatric studies or pathological changes. Registration between two very different anatomies is difficult mainly for two reasons. First, an objective function such as mean-squared error (MSE) or mutual information (MI) is a highly nonlinear function of the deformation variable f, and therefore the optimization process is often trapped in a local minimum. Second, the uncertainty about the one-to-one correspondences in the images makes the optimization problem ill-posed. Although regularization on the final deformation f is commonly used to remedy the situation, the regularization alone cannot guide the variable f to an optimal path of registration. Geodesic registration methods have been proposed to cope with these challenges. In the framework of [1,2], the optimal deformation path $f(x,t)$ is computed by solving the Euler-Lagrange equation:

$$\frac{\partial f(x,t)}{\partial t} = v(f(x,t),t), \quad f(x,0) = x,$$

where v is the velocity field, and $t \in [0,1]$. Among the possible solutions of the ODE, the geodesic path is defined as the 'shortest' path that connects two

* This work was supported by NIH grants R01-MH079938 and R01-AG014971.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 680–687, 2009.

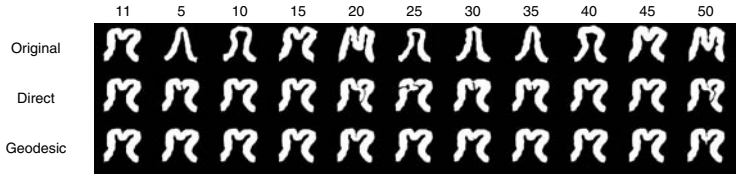


Fig. 1. Examples of synthetic images and typical registration results. The top row shows the subset of original images with the sample numbers on top. The image 11 in the first column is the template. The second and the third rows show the comparison of direct vs the proposed registration methods. Note the unnatural warping of the boundaries in the direct method, especially in image 20, 25, and 50.

images. The length of the geodesic path is defined as a weighted sum of the intensity difference and the smoothness measure of the velocity field v:

$$g^2(I_i, I_j) = \int_\Omega |I_i(x) - I_j(f(x))|^2 \, dx + w \int_0^1 \|v(x,t)\|_L^2 \, dt.$$

When all images are in the single orbit of diffeomorphisms, the quantity g provides the dataset with a true metric structure. Registration methods which are based on this numerical integration of the geodesics provide improvements in the large deformation problems [3,4]. However, the increase in the computational costs can outweigh the benefits of geodesic registration methods for certain problems. More importantly, these methods calculate geodesics on the manifold of diffeomorphic transformations, most of which do not represent valid anatomies. One ideally wants to calculate geodesics on the manifold of all transformations that represent true biological variations, which unfortunately cannot be represented analytically.

We aim to achieve this goal by learning this manifold from the data. In this paper, we approximate the shortest path on the continuous space of diffeomorphisms by the shortest path on a k-nearest-neighbor(kNN) graph that approximates the metric structure of underlying data manifold. A large deformation that maps an anatomy to another, is decomposed into a series of small deformations along the path on the graph. This means that the large deformation actually progresses through a series of small deformations between the anatomies in the data. Since only nearby images need to be registered to each other, a simple registration algorithm suffices to solve the difficult registration problem. We will call this approach a **pseudo-geodesic** registration method, because the registration is restricted to the paths on the assumed manifold of the data.

The advantages of the proposed registration method include the following:

1. **Learning of anatomical manifolds**: Our method computes the geodesics on the manifold of the actually observed anatomical variations from the data, instead of the infinite-dimensional manifold of diffeomorphisms.

2. **Efficient computation**: Our method solves a large deformation problem by using simpler and faster algorithms such as the Diffeomorphic Demons algorithm [5] which we use for pairwise registrations.
3. **Visualization**: Our method provides a low-dimensional characterization of the data structure by flattening the learned manifold.
4. **Automatic template selection**: The optimal template is chosen with a criteria consistent with the pseudo-geodesics, instead of being solved iteratively as in [4,3].
5. **Meta-registration**: Our approach is flexible to the choices of registration cost or an algorithm as its components.

The examples in Figure 1 demonstrate the accuracy of the proposed approach achievable with the same registration method that fails to find optimal solutions when used directly without the pseudo-geodesic paths.

Related Work. The pseudo-geodesic path relies on the learning methods for discovering the empirical manifold of data instead of numerical methods of solving ODEs on an analytic manifold of shapes. In our method we adopt the Isomap algorithm [6] to compute and visualize the manifold structure of data, which was originally proposed for nonlinear dimensionality reduction of high-dimensional data. Several authors have proposed the learning-based methods for template selection and visualization of the data [7,8,9]. Although their goals are different from ours, these methods build the models of data by pairwise registration of multiple images similarly to our method. However, those works lack the counterpart the pseudo-geodesics which is the key to the computational efficiency of the proposed method in this paper.

The remainder of the paper is organized as follows. Section 2 describes the proposed algorithms in detail, Section 3 illustrates the application of the algorithm to 2D and 3D image databases, and Section 4 concludes the paper.

2 Pseudo-geodesic Registration

In this section we provide the algorithmic details of the pseudo-geodesic registration. The overall registration procedure consists of three stages which are described in the following three subsections. First, we analyze the data structure by coarse registrations between all image pairs. From this we find a kNN graph structure and a low-dimensional embedding of the data. In the second stage, we choose a template automatically from the graph structure, and identify the pseudo-geodesic paths from the template to the other samples. In the third stage, we compute the final large deformation by composing small deformations between adjacent images along the paths.

Throughout the paper, let's assume the dataset $\mathcal{I}$ consists of n images $I_1, ..., I_n$, and each image is a nonnegative real function on a 2D or a 3D domain Ω.

2.1 Construction of Empirical Manifolds

In the first stage we construct the empirical manifold of data by investigating its metric structure. For this purpose we represent the data as a graph whose vertices correspond to the image samples. Below is the summary of the required steps.

1. Perform a coarse registrations between all pair of images. The minimum of the registration cost function between I_j is I_i after registration is considered the length d_{ij} of the edge e_{ij}.[1]
2. Construct a connected kNN or ϵ-NN graph based on the edge lengths.
3. Find the pseudo-geodesics (=shortest paths on the graph) between all pairs of vertices, e.g., by Dijkstra's or Floyd-Warshall algorithm. The length g_{ij} of a pseudo-geodesic is the sum of its edge lengths d_{kl} along the path.
4. (Optional) Visualize the Euclidean embedding of the data by solving eigenvalue problems (refer to [6] for details).

One can show that the shortest path length g_{ij} is a valid metric by construction regardless of the properties of the constituent distance d_{ij}. The most time-consuming part in practice is step 1 which requires $O(n^2)$ registrations between all image pairs. To reduce the overhead we perform the registration on the coarse resolution images of the original data, and also use fewer number of iterations than the final registration. This heuristic is justified by the observation that the kNN graph topology does not change much by the approximation.

2.2 Automatic Template Selection

An unbiased template of the given data can be defined as the geodesic mean of the data [4,3]. From the graph derived in the previous section, we can choose a template from the samples that is closest to the pseudo-geodesic mean:

$$I_T = \arg\min_i \sum_j g^2(I_i, I_j).$$

Since we have already computed the geodesic lengths g_{ij}, the template can be chosen by looking-up of the values. Two other alternatives to the mean are the center ($I_T = \arg\min_i \max_j g(I_i, I_j)$) and the median ($I_T = \arg\min_i \sum_j g(I_i, I_j)$) of the graph. The three templates look similar in our experiments. We choose the median as the template due to its resilience to outlier samples in the data.

2.3 Computation of Large Deformations

We compute the large deformation from the template I_T to any node I_j by a recursive composition of the small deformations from its edges along the pseudo-geodesic path. Let $f_{i,j} : \Omega_i \to \Omega_j$ denote the deformation field computed from

[1] The cost function of a generic registration algorithm is a weighted sum of the intensity difference $\int_\Omega |I_i(x) - I_j(f(x))|^2 dx$ and the smoothness measure, e.g., $\int_\Omega \nabla^2 f(x)\, dx$.

the registration of I_j to I_i. Given the two fields $f_{i,j}$ and $f_{j,k}$, we can easily compute the composition field $f_{i,k} = f_{j,k} \cdot f_{i,j} : \ \Omega_i \rightarrow \Omega_k$ by resampling and interpolation of the two fields.[2] The final deformation $\hat{f}_{T,j}$ is the refinement on the composed field $f_{T,j}$ by a few additional iterations of registration. This removes the numerical errors accumulated from the composition of deformations. Below we summarize the procedure.

1. Identify n pseudo-geodesic paths from I_T to the rest I_j, $\forall j \in 1, ..., n$.
2. Enumerate all edges $\mathcal{E}$ used in any of the shortest paths. Perform accurate registration between (I_i, I_j), $\forall e_{ij} \in \mathcal{E}$.
3. For each $j \in 1, ..., n$,
   (a) Let $s = (s_1 = T, ..., s_m = j)$ be the pseudo-geodesic path from I_T to I_j.
   (b) If $f_{T,j}$ is already computed then exit.
   (c) Otherwise, recursively compute $f_{s_1,s_m} = f_{s_{m-1},s_m} \cdot f_{s_1,s_{m-1}}$.
   (d) Fine-tune f_{s_1,s_m} by a few additional iterations of registration.

Note that we needed only coarse registration results in the previous stages, and this stage is where we actually perform accurate registrations. The step 2 may seem to be a huge computational burden at first since the number of all the edges in a graph can be a large as n^2. In fact, we only need to update the registration for $n - 1$ edges, that is, no more than the number of direct registration for a conventional approach. This is due to property of the graph that the shortest paths from the template to the rest forms a spanning tree. Furthermore the registration converges faster since the two adjacent images are similar by construction.

2.4 Discussion

The proposed registration method is motivated by the Isomap algorithm, which proves that a true geodesic on a convex set can be approximated by a path on the kNN graph connecting the data samples. Due to the finite number of samples, the pseudo-geodesic is different from the numerically integrated geodesic in general. Moreover, the different choices of the neighborhood size k in kNN yield different graphs and therefore different paths. However, we can show that the final large deformations are robust to the these approximations: two different pseudo-geodesic paths between the common image pair result in the same final deformation as long as each step of the paths is smooth enough to avoid any local minimum of registration.

3 Experiments

3.1 Synthetic Images

The data consist of 60 binary images of size 140×140 which simulate a patch of a cortex varying in shape and the number of folds. We use an ITK [10] version

[2] Two deformations should be properly composed, not simply added to each other.

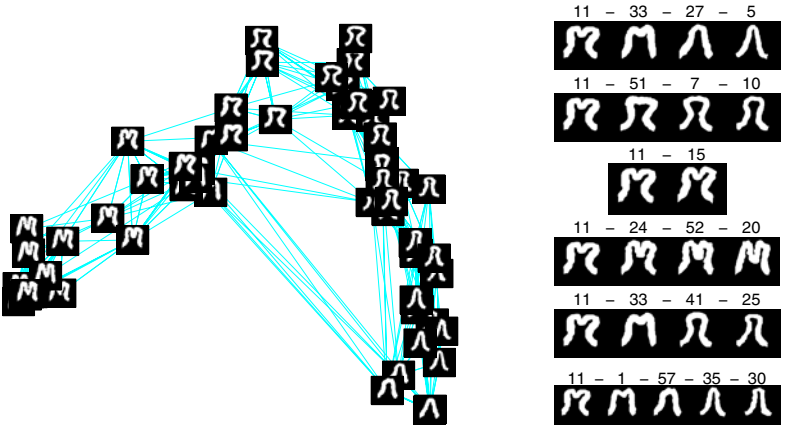


Fig. 2. Left. Two-dimensional embeddings of the manifold of synthetic shapes. The embedding reveals that there are two or three major variants (which resemble M, Ω, Λ) and the rest of the images can be interpolated from those representative image. **Right**. Examples of the pseudo-geodesic paths from the template 11 to 5, 10, 15, 20, 25, and 30. Note the gradual and monotonic change of shapes along each path.

Table 1. Final registration error of the synthetic images : comparison of Mean Square Error (MSE), Harmonic Energy (HE), and Maximum of Jacobian Determinant(MJD). The (avg) and (max) are the average and maximum of all images in the data.

Method	MSE (avg)	MSE (max)	HE (avg)	HE (max)	MJD (avg)	MJD (max)
Direct	388	1136	0.260	0.583	4.01	14.6
Pseudo-geodesic	327	565	0.218	0.504	3.55	7.70

of the Diffeomorphic Demons [5] for registration due to its efficiency. Images are registered with three levels of resolution ($10 \times 10 \times 10$) and the smoothing kernel size of 1.5. Figure 2 shows the two-dimensional Euclidean embedding of the synthetic database and a few examples of pseudo-geodesic paths. The registration results are shown in Figure 1. For comparison we perform the direct registration of all images to the template without following the pseudo-geodesic paths. One can visually confirm the superior quality of the pseudo-geodesic registration results over the direct registration results obtained from the same algorithms and parameters. Table 1 summarizes the final registration errors. For each of the six error criteria, the pseudo-geodesic method achieves smaller error than the direct registration method does.

3.2 Fractional Anisotropy Map of Mouse

Next we apply our registration method to a dataset of mouse brains which is collected with the aim of creating a normative atlas of a developing mouse brain. The data consist of 69 Fractional Anisotropy volume of the brains sampled at 2,3,4,7,10,15,20,30,45, and 80 days of age. Each volume is resized to $150 \times 150 \times 100$

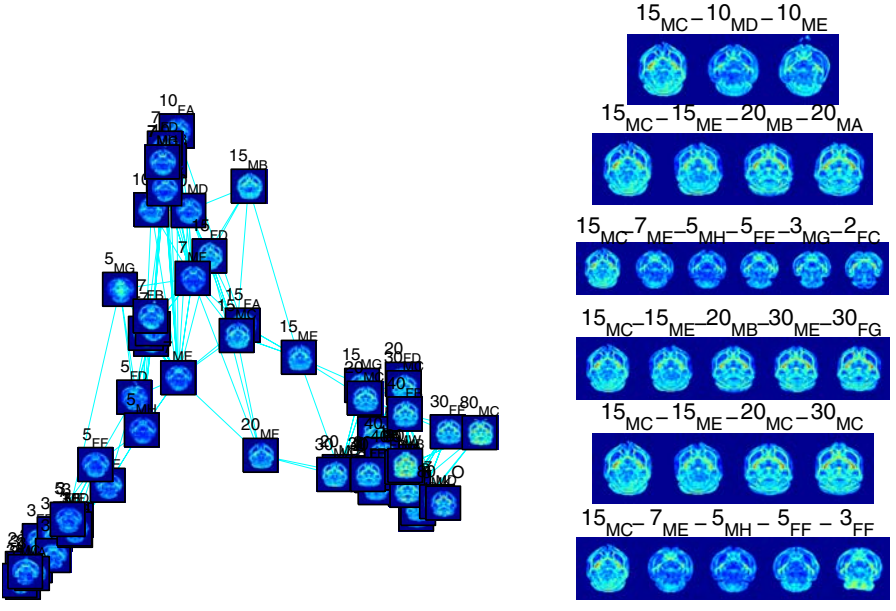


Fig. 3. Left. Two-dimensional embeddings of the manifold of mouse brains. A mid-axial slice is shown for each mouse brain volume. The number on top of each image is the age and name of the sample. **Right**. Examples of the pseudo-geodesic paths from the template to others. Each path reflects the changes in both the shapes and the appearances of developing brains.

Table 2. Final registration error of the mouse brain images. See Table 1 for notations.

Method	MSE (avg)	MSE (max)	HE (avg)	HE (max)	MJD (avg)	MJD (max)
Direct	19.33	40.11	0.0984	0.1927	9.051	19.92
Pseudo-geodesic	19.98	41.62	0.0800	0.1345	8.789	13.02

and affinely aligned. We also use the Diffeomorphic Demons method for pairwise registration, with two levels of resolution 15×10 and with $\sigma = 1.0$. The images in this dataset not only have a larger number of voxels than the synthetic data, but they are more challenging for registration due to their large shape and appearance variations from different ages and the degrees of maturation of tracts.

The two-dimensional embedding of the data in Figure 3 (left) provides a glimpse of its manifold structure. From the figure we can observe that the major variability of the data comes from the age. The importance of the age factor is also observed in Figure 3 (right): a path that connects two brain images of different ages passes through brains of intermediate ages in a monotonic fashion. These findings are consistent with our prior knowledge of the database. Table 2 summarizes the final registration errors. The MSE of the pseudo-geodesic method is slightly worse but comparable with the MSE of the direct method. However,

the harmonic energy and the Jacobian determinants are significantly lower in the pseudo-geodesic method. One possible interpretation is that the pseudo-geodesic method can achieve the similar level of intensity difference as the direct method but with a much smoother deformation.

4 Conclusion

In this paper we propose a novel method of performing a large deformation registration. The most distinguishing feature of the method is that it computes the geodesics on the manifold of the observed anatomical variations from the data, instead of computing the geodesics by solving ODEs. Furthermore, the learned manifold provides a visualization of the data structure and allows us to choose an optimal template among the samples. We describe an meta-registration algorithm to efficiently compute large deformations, which can use a large class of registration algorithms as a component. The experiments have their caveats but show promising results. We are currently testing the algorithms on a number of other image databases. A comprehensive study will be reported in the near future.

References

1. Grenander, U., Miller, M.I.: Computational anatomy: an emerging discipline. Quart of Appl. Math., 617–694 (1998)
2. Miller, M., Trouve, A., Younes, L.: On the metrics and euler-lagrange equations of computational anatomy. Ann. Rev. of Biomed. Eng. 4, 375–405
3. Avants, B., Gee, J.C.: Geodesic estimation for large deformation anatomical shape averaging and interpolation. NeuroImage 23, S139–S150 (2004)
4. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. Neuroimage (23 suppl. 1) (2004)
5. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Non-parametric diffeomorphic image registration with the demons algorithm. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 319–326. Springer, Heidelberg (2007)
6. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319–2323 (2000)
7. Blezek, D.J., Miller, J.V.: Atlas stratification. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 712–719. Springer, Heidelberg (2006)
8. Sabuncu, M.R., Balci, S.K., Golland, P.: Discovering modes of an image population through mixture modeling. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 381–389. Springer, Heidelberg (2008)
9. Rohde, G.K., Ribeiro, A.J.S., Dahl, K.N., Murphy, R.F.: Deformation-based nuclear morphometry: Capturing nuclear shape variation in hela cells. Cytometry Part A 73A(4), 341–350 (2008)
10. Ibanez, L., Schroeder, W., Ng, L., Cates, J.: The ITK Software Guide. Kitware, Inc. (2005), http://www.itk.org/ItkSoftwareGuide.pdf

A Non-rigid Registration Method for Serial microCT Mouse Hindlimb Images

Jung W. Suh[1], Dustin Scheinost[1], Donald P. Dione[2], Lawrence W. Dobrucki[2], Albert J. Sinusas[1,2], and Xenophon Papademetris[1,3]

[1] Department of Diagnostic Radiology, Yale School of Medicine, New Haven, CT
[2] Department of Internal Medicine, Yale School of Medicine, New Haven, CT
[3] Department of Biomedical Engineering, Yale University New Haven, CT

Abstract. We present a new method for the non-rigid registration of serial mouse microCT images which undergo potentially large changes in the positions of the legs due to articulation. While non-rigid registration methods have been extensively used in the evaluation of individual organs, application in whole body imaging has been limited, primarily because the scale of possible displacements and deformations is large resulting in poor convergence of most methods. Our method is based on the extended demons algorithm that uses a level-set representation of the mouse skin and skeleton as an input, and composed of three steps reflecting the natural physical movements of bony structures. We applied our method to the registration of serial microCT mouse images demonstrating encouraging performances as compared to competitive techniques.

1 Introduction

Whole body mouse registration is a challenging emerging problem in image analysis with many potential applications. These applications arise from the extensive genetic similarities between the mouse and human genomes which make the mouse an excellent model for understanding human developmental processes and pathologies. A particular mouse model of interest in our work is that of hindlimb ischemia [1], which is used to study vascular remodeling. In this model an ischemic zone is created in one limb by surgical ligation of the femoral artery, with the other limb serving as a paired control. Vascular remodeling is an important process which plays an important role in both peripheral arterial disease and coronary artery disease.

Molecular imaging using microSPECT can be used to quantify angiogenesis, the process of formation of new blood vessels. This is often performed using hybrid microSPECT/microCT scanners. A key challenging in the quantification of the SPECT images is the parcellation of the hindlimbs into different anatomical regions (e.g. proximal, distal and foot) to enable the region-of-interest analysis of these images. Whereas such parcellation can be done manually (using the anatomy from the co-aquired microCT), this is tedious and leads to subjective divisions. Our goal in this work is to develop atlas-based labeling of the hindlimbs where the parcellation is done manually at the first time point and automatically

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 688–695, 2009.

propagated in time using our non-rigid registration method which we describe in this paper.

Registering the hindlimbs is probably the most challenging aspect of whole body mouse registration. Previous approaches to the problem of whole body mouse registration [6,7,8] have used a combination of the robust point matching method (RPM) to first align the bone surface as an initial condition followed by an intensity-based registration which aims to improve accuracy in the soft-tissue structures. Our work differs from the previous work in the literature in that (i) we adopt an implicit surface registration method which relies on a distance-map parameterization of the bone and skin surfaces and (ii) in performing the surface registration we follow the natural "anatomical" constraints of first registering the bone surfaces – which literally anchor the rest of the body and then refining this by introducing skin surface information. The result of the surface-based registration is then augmented by an intensity based non-rigid registration as before. We validate our results using 14 image-pairs using both intensity similarity criteria and measures of (implicit) segmentation accuracy as evaluated using standard overlap measures.

We compare our results to standard methods such as the demons method (which uses the same transformation model and optimization method as our method), the intensity-based non-rigid registration with free-form deformation (IFFD)[11] and a robust point-based method (RPM)[9] – which forms the basis of most of the previous mouse non-rigid registration work in the literature.

2 Methods

Figure 1 shows the brief block diagram of the proposed method. The inputs are microCT images of the same mouse imaged over a period of 4 weeks. This method is a three-step serial registration composed of a bone surface registration, a skin surface registration and an intensity registration. After initially executing the bone surface registration(f_1), we execute the skin surface registration(f_2) based on the initial bone surface registration result. Then, we refine the image intensities for the soft tissue between the bone and skin by the final intensity registration(f_3). We combine these three registrations as one registration using concatenation ($f_{total} = f_3 \circ f_2 \circ f_1$). We first review an extended demons method for single surface mapping such as a bone surface. Then we describe each step in the proposed method.

2.1 Extended Demons Registration

Thirion [2] suggested an image-matching algorithm using a diffusion process. Using the demons concept from Maxwell's demons in thermodynamics, the non-rigid image matching process is modeled as a diffusion process in the demons algorithm. The theory behind this algorithm is that the object boundary can change its shape depending on the position of the demons within the image domain. The demons' forces, which deform the object shape, are generated to reduce the disparity between the reference and the moving images. Equation (1)

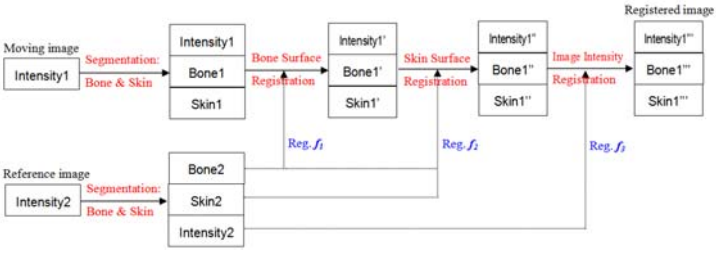


Fig. 1. Block diagram for our registration method

shows the demons algorithm based on the optical flow equation. Here, $\mathbf{r}(\mathbf{X})$ is the reference image, $\mathbf{m}(\mathbf{X})$ is the moving image and $\mathbf{v}(\mathbf{X})$ is the displacement.

$$\mathbf{v}(\mathbf{X}) = -((\mathbf{m}(\mathbf{X}) - \mathbf{r}(\mathbf{X}))\nabla\mathbf{r}(\mathbf{X}))/(\|\nabla\mathbf{r}(\mathbf{X})\|^2 + (\mathbf{m}(\mathbf{X}) - \mathbf{r}(\mathbf{X}))^2) \tag{1}$$

Many studies improved the original demons algorithm by increasing the speed of convergence [4,5]. For stronger convergence performance, the extended demons registration method takes advantage of the level-set representation of the object boundary using a distance map based on the demons algorithm. The extended demons algorithm utilizes the distance map around the shape border instead of intensity which results in improved convergence even though the two objects are not overlapped sufficiently. Thus, the extended demons registration algorithm can achieve much higher convergence speed with better confidence in binary volume registration. In this paper, we used a symmetric version of the extended demons method to avoid local minima as below.

$$\mathbf{v}(\mathbf{m}, \mathbf{r}, \mathbf{X}) = -\frac{2(\mathbf{m}(\mathbf{X}) - \mathbf{r}(\mathbf{X}))(\nabla\mathbf{r}(\mathbf{X}) + \nabla\mathbf{m}(\mathbf{X}))}{\|\nabla\mathbf{r}(\mathbf{X}) + \nabla\mathbf{m}(\mathbf{X})\|^2 + (\mathbf{m}(\mathbf{X}) - \mathbf{r}(\mathbf{X}))^2} \tag{2}$$

2.2 First Step: Registration of Bone Surface

From each reference and moving CT images of the same mouse, binary bone images and binary skin images are segmented through intensity differences by thresholding and clustering (K-means clustering with 3 classes)(Figure 1). The level-set represented bone images are generated by distance map from the binary reference bone image and the binary moving bone image respectively. Using $\mathbf{v}(\mathbf{m}_B, \mathbf{r}_B, \mathbf{X})$ of the extended demons algorithm (equation (2)), the deformation between the two level-set represented bone images is calculated.

Here, $\mathbf{r}_B$ is the reference image represented by level-set from the bone boundary, $\mathbf{m}_B$ is the moving image represented by level-set from the bone boundary, and $\mathbf{v}(\mathbf{m}_B, \mathbf{r}_B, \mathbf{X})$ is the displacement between $\mathbf{r}_B$ and $\mathbf{m}_B$.

2.3 Second Step: Registration of Skin Surface

After the registration of bone surface in the first step, the moving skin image is deformed with the displacement of the bone surface registration result. In the same way as the bone registration, the level-set represented skin images

are generated by distance maps from the binary reference skin image and the binary moving skin image respectively. Using $\mathbf{v}(\mathbf{m}_{1S}, \mathbf{r}_S, \mathbf{X})$, we calcualte the deformation between the two level-set represented skin images. Here, $\mathbf{r}_S$ is the reference image represented by level-set from the skin boundary, $\mathbf{m}_{1S}$ is the moving image represented by level-set from the skin boundary (deformed by the result $(\mathbf{v}(\mathbf{m}_B, \mathbf{r}_B, \mathbf{X}))$ of the first step), and $\mathbf{v}(\mathbf{m}_{1S}, \mathbf{r}_S, \mathbf{X})$ is the displacement between $\mathbf{r}_S$ and $\mathbf{m}_{1S}$. In this second step, bone surfaces are registered again simultaneous (during a same iteration) with the skin surface registration using $\mathbf{v}(\mathbf{m}_{1B}, \mathbf{r}_B, \mathbf{X})$. Here, $\mathbf{m}_{1B}$ is the moving image represented by level-set from the bone boundary (deformed by the result $(\mathbf{v}(\mathbf{m}_B, \mathbf{r}_B, \mathbf{X}))$ of the first step), and $\mathbf{v}(\mathbf{m}_{1B}, \mathbf{r}_B, \mathbf{X})$ is the displacement between $\mathbf{r}_B$ and $\mathbf{m}_{1B}$.

During each iteration of the extended demons algorithm the two displacements $(\mathbf{v}(\mathbf{m}_{1B}, \mathbf{r}_B, \mathbf{X})$ and $\mathbf{v}(\mathbf{m}_{1S}, \mathbf{r}_S, \mathbf{X}))$ are spatially combined in every iteration by equation (3) within the f_2 in Figure 1.

$$\mathbf{v}(\mathbf{X}) = \mathbf{w}_B(\mathbf{X})\mathbf{v}(\mathbf{m}_{1B}, \mathbf{r}_B, \mathbf{X}) + \mathbf{w}_S(\mathbf{X})\mathbf{v}(\mathbf{m}_{1S}, \mathbf{r}_S, \mathbf{X}) \tag{3}$$

Where $\mathbf{w}_B(\mathbf{X})$ and $\mathbf{w}_S(\mathbf{X})$ are weight factors defined as:

$$w_B(x_k) = \begin{cases} 1 & \text{if } d_B(x_k) \le 0 \text{ and } d_S(x_k) < 0 \\ 0 & \text{if } d_B(x_k) > 0 \text{ and } d_S(x_k) \ge 0 \\ \frac{|d_S(x_k)|}{|d_B(x_k)|+|d_S(x_k)|} & \text{if } d_B(x_k) > 0 \text{ and } d_S(x_k) < 0 \end{cases} \tag{4}$$

$$w_S(x_k) = \begin{cases} 0 & \text{if } d_B(x_k) \le 0 \text{ and } d_S(x_k) < 0 \\ 1 & \text{if } d_B(x_k) > 0 \text{ and } d_S(x_k) \ge 0 \\ \frac{|d_B(x_k)|}{|d_B(x_k)|+|d_S(x_k)|} & \text{if } d_B(x_k) > 0 \text{ and } d_S(x_k) < 0 \end{cases} \tag{5}$$

Here $|d_B(x_k)|$ is the absolute value of the signed distance value and k is the voxel index. The displacement of the bone registration $\mathbf{v}(\mathbf{m}_{1B}, \mathbf{r}_B, \mathbf{X})$ is exclusively used within the bone region, and the displacement of the skin registration $\mathbf{v}(\mathbf{m}_{1S}, \mathbf{r}_S, \mathbf{X})$ is exclusively used outside the skin region. The displacement of the soft tissue between bone and skin is calculated by interpolation depending on the distance from both the bone and skin. In order to reduce the abrupt change of the displacement, the standard deviation of the smoothing filter within the extended demons algorithm is set to a large value ($\sigma = 2$ voxels.)

2.4 Third Step: Registration of Graylevel Intensity Image

After finishing the registration of bone and skin surfaces, soft tissues between the bone and skin are registered through this third step. In the third step, the symmetric version of the demons algorithm [2] with histogram matching is used for the image intensity registration instead of the extended demons algorithm. The moving graylevel intensity image is deformed with both the displacement (f_1) of the bone surface registration result and the displacement (f_2) of the skin surface registration result as shown in Figure 1. The deformed moving graylevel intensity image and the reference graylevel intensity image are used as inputs of this third step of registration.

2.5 Implementation Details

Three resolution layers were utilized in each step from the coarsest to the finest layer to search the displacement in the 3D data set. Iteration numbers for each layer were set to 200, 200 and 30 from the coarsest to the finest layer, respectively.

The images consisted of the lower half of the mouse (roughly from below the lungs to the feet). The image size after cropping was approximately 160 × 90 × 140 (voxels).

3 Results

We present results of the application of our method to $N = 14$ pairs of mouse microCT images, acquired with a resolution of 100 × 100 × 100 μm^3. In these images the mice are virtually divided into 7 regions shown in Figure 2. These enables the comparison of the overlap of these regions after registration. In order to generate this mouse parcellation, we manually positioned planes to segment the

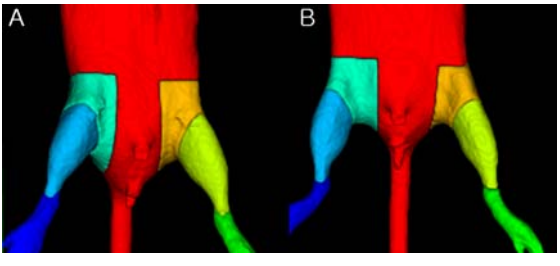


Fig. 2. 7-region mouse parcellation; body and tail (red), the left proximal hindlimb (yellow), the left distal hindlimb (light green), the left foot (green), the right proximal (light teal) , the right distal (light blue), and the right foot (blue) (a) Baseline, (b) Moving

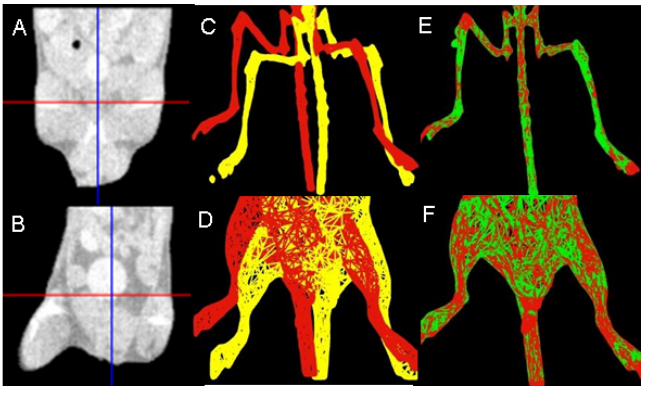


Fig. 3. (A) and (B) Reference and Targed Images. Bone (C) and Skin (D) Surfaces prior to Registration. Bone (E) and Skin (F) Surfaces after registration.

Table 1. Similarity comparisons (Jaccard Similarity Coefficient) for the regions

Right proximal	*JSC* Averg.	T test
First step	67.66	1.65E-04
Second step	78.36	0.66
Third step (Final reg)	**78.92**	
IFFD	76.50	0.30
RPM	73.14	0.02
Demons	64.67	1.10E-02

Left proximal	*JSC* Averg.	T test
First step	65.31	3.9E-06
Second step	75.135	0.44
Third step (Final reg)	**75.70**	
IFFD	72.56	6.65E-05
RPM	74.26	0.29
Demons	67.71	1.53E-02

Right distal	*JSC* Averg.	T test
First step	63.44	2.03E-05
Second step	82.60	0.19
Third step (Final reg)	**85.26**	
IFFD	85.46	0.85
RPM	79.49	9.08E-04
Demons	53.96	1.73E-03

Left distal	*JSC* Averg.	T test
First step	62.40	9.76E-07
Second step	79.45	0.31
Third step (Final reg)	**80.42**	
IFFD	76.79	5.03E-02
RPM	70.78	0.029
Demons	59.52	1.70E-03

Right foot	*JSC* Averg.	T test
First step	60.53	4.61E-04
Second step	79.03	0.33
Third step (Final reg)	**83.29**	
IFFD	60.51	3.09E-04
RPM	58.77	6.53E-08
Demons	34.34	8.39E-05

Left foot	*JSC* Averg.	T test
First step	64.92	5.21E-08
Second step	78.72	0.40
Third step (Final reg)	**77.50**	
IFFD	55.36	7.54E-06
RPM	47.81	1.45E-06
Demons	19.57	2.13E-06

mouse into 7 regions; main body including the tail, the left proximal hindlimb, the left distal hindlimb, the left foot, the right proximal hindlimb, the right distal hindlimb, and the right foot. The planes were positioned on a 3D rendering of the CT image. Among those 7 regions, the main part of the body including the tail, is excluded as its upper boundary is variable between images. We compared our method to two non-rigid registration methods ([11,9]) and to the demons algorithm. We demonstrate encouraging improvement in registration accuracy compared to these standard methods. As a similarity measure for local shape, we used the Jaccard similarity coefficient ($JSC = (\mid R \cap M \mid / \mid R \cup M \mid) \times 100$). Where M is the moving and R is the reference object, and $|M|$ refers the number of members in the set M as a cardinality.

Table 1 shows the comparison results with the Jaccard Similarity Coefficient (JSC) for the six mouse regions. In the first four tables in Table 1, the average JSCs of our proposed method (Third step) are slightly better or similar (Right distal hindlimb) to the average JSCs of IFFD, and always better than those of the Demons algorithm and RPM. In cases of the JSCs for both the right foot and left foot, however, our proposed serial registration has much

better than all methods (much less than 0.05 in paired-T test). We note that our method achieves consistently high overlaps over the all six mouse regions and the improvement is particularly noticeable in the most challenging part – the feet. This is probably due to two reasons. The intensity based only methods (IFFD,demons) can not converge in these region due to the large displacements that can be present. While the imaging technologists attempt to place the mice in roughly the same orientation (the scans are acquired in weekly intervals) – this is almost impossible in the case of the feet resulting in large displacements which are less easily estimated by intensity-based methods. In the case of RPM, the improvement is most likely caused by the fact that the number of points that can be used for RPM is finite (we used about 5000 points) as it requires the computation of a match matrix which even if sparse, still becomes computationally intractable. This results in less accurate surface representation of smaller structures which explains partially why our method becomes increasingly superior to RPM as we move down the limb. (Please note that the surface only part of our algorithm – Second Step – also gives superior results to RPM, a surface-only technique, which is a more fair comparison.)

4 Discussion and Conclusions

We presented a new non-rigid stepped registration method for quantitiative evaluation of serial mouse images. Our method, which is based on an extended demons formulation, first aligns the bone surfaces to achieve a good initial condition and then leverages additional information obtained from skin surfaces to improve the result and finally includes the use of direct intensity information constrained by the pre-computed surface-derived transformations maps. These first and second steps of surface mappings are critical in ensuring that the intensity information is useful, as without a good initialization, intensity only methods fail to converge to the appropriate result. We evaluated the proposed method on 3D volumes from 14 serial mouse microCT data sets, and demonstrate that our method performs better three standard methods (IFFD, Demons and RPM.) The first two methods represent popular standard nonrigid intensity based registration algorithms which are readily available – and hence should be tried on a new problem prior to developing a customized method. The comparison to RPM was done as this has been used previously for whole body mouse registration.

Relating the comparison with Li *et al.* [7,8], this method is also a two-step algorithm. Step one is bone registration using robust point matching (RPM) from bone surface; step two is intensity-based registration. Although we don't use the same intensity-based registration method in the final step as Li *et al.*, the intensity-based registrations in both cases are a "finishing" step. While we were not able to compare our results directly to the work of Li *et al.*, we note that the surface only part of this algorithm (i.e. Step 2) demonstrates superior performance to the surface only portion of that method (which relied on RPM) for both sets of criteria. This is a strong indication that all things being equal our combined method is likely to perform better than this competitor in the challenging area of the hindlimbs.

Our methodology may also be applicable to serial imaging of patients with peripheral arterial disease. In addition this type of multistage registration may be applicable to registration problems in the abdomen where there is the potential for large displacements of key structures such that conventional intensity-based methods are not able to handle. However, we emphasize that in much of this work, the goal is *not* to use the mouse as a test-bed for the development of imaging/image analysis methods which will eventually be translated to human applications. The goal is, rather, to *develop* mouse specific methods that will help understand the underlying pathology. It is *this increased understanding* that will ultimately most likely prove useful in a translational sense, as opposed to the methods themselves.

Acknowledgments. This work was supported in part by the NIH under grants R01HL065662 (AJS) and R01EB006494 (XP).

References

1. Nikol, S., Huehns, T.Y.: Preclinical and clinical experience in vascular gene therapy: advantages over conservative/standard therapy. Journal of Invasive Cardiology 13(4), 333–338 (2001)
2. Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis 2(3), 243–260 (1998)
3. Papademetris, X., Dione, D.P., Dobrucki, L.W., Staib, L.H., Sinusas, A.J.: Articulated rigid registration for serial lower-limb mouse imaging. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 919–926. Springer, Heidelberg (2005)
4. Wang, H., Dong, L., O'Daniel, J., Mohan, R., Garden, A.S., Ang, K.K., Kuban, D.A., Bonnen, M., Chang, J.Y., Cheung, R.: Validation of an accelerated 'demons' algorithm for deformable image registration in radiation therapy. Physics in Medicine and Biology 50, 2887–2905 (2005)
5. Cachier, P., Pennec, X., Ayache, N.: Fast Non Rigid Matching by Gradient Descent: Study and Improvements of the "Demons" Algorithm. Rapports de recherche no 3706, Note 25 (1999)
6. Papademetris, X., Shkarin, P., Staib, L.H., Behar, K.L.: Regional Whole Body Fat Quantification in Mice. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 369–380. Springer, Heidelberg (2005)
7. Li, X., Peterson, T.E., Gore, J.C., Dawant, B.M.: Automatic Registration of Whole Body Serial Micro CT Images with a Combination of Point-based and Intensity-based Registration Techniques. In: ISBI 2006, pp. 454–457 (2006)
8. Li, X., Yankeelov, T.E., Peterson, T.E., Gore, J.C., Dawant, B.M.: Automatic nonrigid Registration of Whole Body CT Mice Images. Medical Physics 35(4), 1507–1520 (2008)
9. Chui, H., Rangarajan, A.: A new point matching algorithm for nonrigid registration. Comput. Vis. Image Underst. 89, 114–141 (2003)
10. Martin-Fernandez, M.A., Munyoz-Moreno, E., Martin, M.: Articulated registration: elastic registration based on a wire-model. In: Proceedings of SPIE 2005, vol. 5747, pp. 182–191 (2005)
11. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Hawkes, D.J.: Non-rigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging 18, 712–721 (1999)

Non-rigid Image Registration with Uniform Gradient Spherical Patterns

Shu Liao and Albert C.S. Chung

Lo Kwee-Seong Medical Image Analysis Laboratory,
Department of Computer Science and Engineering,
The Hong Kong University of Science and Technology, Hong Kong
liaoshu@cse.ust.hk, achung@cse.ust.hk

Abstract. In this paper, we propose a new feature based non-rigid image registration method for dealing with two important issues. First, in order to establish reliable anatomical correspondence between template and subject images, efficient and distinctive region descriptor is needed as intensity information alone maybe insufficient. Second, since interference factors such as monotonic gray-level bias fields are commonly existed during the imaging process, the registration algorithm should be robust against such factors. There are two main contributions presented in this paper. (1) A new region descriptor, named uniform gradient spherical pattern (UGSP), is proposed to extract the geometric features from input images. UGSP encodes second order voxel interaction information. (2) The UGSP feature is rotation and monotonic gray-level bias field invariant. The proposed method is integrated with the Markov random field (MRF) labeling framework to formulate the registration process. The α-expansion algorithm is used to optimize the corresponding MRF energy function. The proposed method is evaluated on both the simulated and real 3D databases obtained from BrainWeb and IBSR respectively and compared with other state-of-the-art registration methods. Experimental results show that the proposed method gives the highest registration accuracy among all the compared methods on both databases.

1 Introduction

Non-rigid image registration is an active research topic during the last decade. It plays an important role in group analysis, image-guided surgery, atlas superposition and etc. During the last decade, many novel methods were proposed to tackle the non-rigid image registration problem. They can be generally classified into three categories: (1) Landmark based methods; (2) Intensity based methods and (3) Feature based methods. Landmark based registration methods [1,2] first select a set of landmark points from template and subject images. Then the optimum transformation is estimated based on the features extracted from those landmark points. In order to produce accurate registration results, a large number of landmark points are required, which brings additional computation burdens. Intensity based registration methods are usually fully automatic. They define similarity measure metrics based on the intensity distributions of input images to guide

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 696–704, 2009.

registration [3,4]. However, intensity similarity does not necessarily equivalent to anatomical similarity. Feature based registration methods use feature vectors to represent each voxel and the registration process is modeled as a feature matching and optimization problem. HAMMER [5] is a representative of this class of methods. The registration accuracy heavily depends on the feature vectors used.

Though the non-rigid image registration problem has been intensively studied during the last decade, it still remains as a challenging task. More precisely, two important issues arise in recent years: First, using intensity information only to characterize anatomical properties of brain images may be insufficient as analyzed in [6]. Therefore, new anatomical descriptor is needed to capture the geometric properties of different anatomical structures. Second, the registration approach should be robust against monotonic gray-level bias fields. Otherwise the algorithm may prefer to align the bias fields between two images instead of aligning their anatomical structures as stated in [7].

To this end, a new feature based non-rigid image registration method is proposed in this paper to deal with these two issues. The major contributions of this paper lie in the following aspects. First, a new region descriptor, named uniform gradient spherical pattern (UGSP), is designed as signatures for each voxel. UGSP encodes second order voxel interaction information. Second, UGSP is rotation and monotonic gray-level bias fields invariant. The UGSP feature is integrated with the Markov random field (MRF) labeling framework to formulate the registration process in this paper. The α-expansion algorithm is used to optimize the corresponding energy function. The proposed method is evaluated on both the simulated and read 3D datasets obtained from BrainWeb and IBSR respectively. Experimental results show that the proposed method achieves the highest registration accuracy among all the compared methods.

2 Uniform Gradient Spherical Patterns

In this section, we describe the new region descriptor, called uniform gradient spherical patterns (UGSP), which is derived based on the local binary patterns (LBP) [8], analyze its properties and show how to use it to represent each voxel as signatures.

Suppose for an input image G. For each voxel $v_c \in G$ a spherical neighborhood is defined centered at v_c with radius R. Also, N samples are taken on the surface of the sphere by using the sampling method proposed in [9], denote them as v_i (i=1,2,...,N). For v_i (i=1,2,...,N) which does not exactly fall in the 3D image grid its intensity is interpolated by tri-linear interpolation. Let ∇v_i denote the gradient of v_i. Then a rotation invariant gradient orientation measure of each neighboring voxel v_i is defined by Equation 1, which is the angle between ∇v_i and $v_c - v_i$, denote as θ_{v_i}.

$$\theta_{v_i} = \arccos \frac{\nabla v_i \cdot (v_c - v_i)}{|\nabla v_i| \cdot |v_c - v_i|}, \tag{1}$$

where $|\nabla v_i|$ and $|v_c - v_i|$ denote the magnitudes of ∇v_i and $v_c - v_i$ respectively.

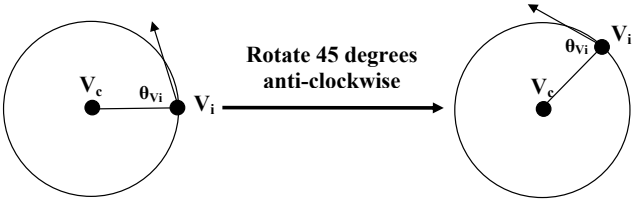


Fig. 1. A 2D example demonstrating the rotation invariant property of θ_{v_i}. v_c is the center voxel, v_i is a neighboring voxel. The black arrow pointing out from v_i denotes the gradient vector of v_i. It is shown that the θ_{v_i} remains the same before and after rotating 45 degrees anti-clockwise.

θ_{v_i} is rotation invariant as it is the angle of the gradient of v_i relative to the direction of $v_c - v_i$. Figure 1 shows a 2D example of θ_{v_i} before and after rotation of 45 degrees. It is shown that θ_{v_i} remains the same.

Then, the space of θ_{v_i} is uniformly partitioned into four subspaces. Each voxel is assigned with a label based on which subspace of θ_{v_i} it belongs. That is, each neighboring voxel is assigned with a label according to Equation 2:

$$l(v_i) = \begin{cases} 1, & if \quad \theta_{v_i} \in [0, \frac{\pi}{4}), \\ 2, & if \quad \theta_{v_i} \in [\frac{\pi}{4}, \frac{\pi}{2}), \\ 3, & if \quad \theta_{v_i} \in [\frac{\pi}{2}, \frac{3\pi}{4}), \\ 4, & if \quad \theta_{v_i} \in [\frac{3\pi}{4}, \pi]. \end{cases} \tag{2}$$

Then the basic gradient spherical pattern (BGSP) is defined based on Equations 1 and 2:

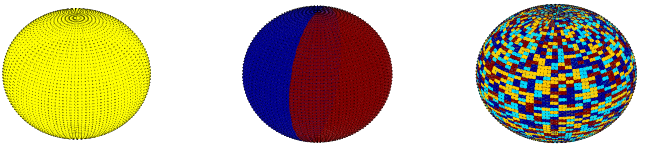
Definition 1. *Basic gradient spherical pattern (BGSP) is the labeled spherical surface obtained from the original spherical neighborhood centered at the reference voxel by using Equations 1 and 2.*

Besides rotation invariance, BGSP is also monotonic gray-level transformation invariant. As monotonic gray-level transformations affect the gradient magnitude of each voxel, the orientation of the gradient remains the same. Therefore, the angle θ_{v_i} between the directions of the gradient vector of v_i and the vector $v_c - v_i$ also remains the same.

However, some of the BGSPs' occurrences are too sparse to reliably reflect the geometric features of input images. Therefore, a subset of BGSPs which represent the fundamental image structures, named uniform gradient spherical patterns (UGSP), are used as signatures of each voxel. Before we define UGSP, we first give the definition of "uniform region" as follows:

Definition 2. *Uniform region is the area on the surface of BGSP where all the voxels belong to this area have the same label defined in Equation 2.*

Then, UGSP is defined as follows:



(a) (b) (c)

Fig. 2. (a) An example of UGSP, its surface can be covered by only one uniform region; (b) An example of UGSP, its surface can be covered by two uniform regions; (c) An example of non-UGSP, its surface cannot be covered by less than or equal to two uniform regions. Different colors represent different labels of each voxel.

Algorithm 1. Determining whether a BGSP is a UGSP

Input: A BGSP with radius R and N neighboring voxels on the surface.

Output: true or false (whether the input is a UGSP or not).

1. Initialize all the neighboring voxels as unflagged.
2. FOR j = 1 to 2
3. IF all the neighboring voxels are flagged
4. Break;
5. ElSE
6. Randomly select an unflagged neighboring voxel v_i on the surface of the input BGSP.
7. Find the largest connected component C_j starting from v_i based on its label $l(v_i)$ using the Breadth First Search (BFS), set a flag for each voxel belonging to that largest connected component. Two neighboring voxels v_m and v_n are considered to be connected if and only if $l(v_m) == l(v_n)$.
8. END IF
9. END FOR
10. IF all the neighboring voxels are flagged
11. Return true
12. ElSE
13. Return false
14. END IF

Definition 3. *Uniform gradient spherical patterns (UGSPs) are BGSPs whose surfaces can be covered by AT MOST two uniform regions.*
For instance, Figure 2(a) is a UGSP as its surface can be covered by only one uniform region. Figure 2(b) is also a UGSP as its surface can be covered by two uniform regions. Figure 2(c) is not a UGSP as its surface cannot be covered by less than or equal to two uniform regions.

UGSP has physical meaning to mirror fundamental image structures. For example, Figure 2(a) denotes that there is an edge at a specific direction as all the neighboring voxels' gradient orientation measures are the same. Figure 2(b) represents that there is a corner formed by two edges. In this paper, all the non-uniform gradient spherical patterns are treated as a single type of pattern. Algorithm 1 is the procedure to determine whether a BGSP is a UGSP or not.

UGSP encodes second order information. The gradient orientation measure defined in Equation 1 already contains the first order interaction information

Algorithm 2. Calculate the UGSP Feature for each Voxel

Input: An input image G, a local cubic square window W for each voxel, the UGSP radius R and the number of neighboring samples N.

Output: A vector image K, each voxel is represented by a UGSP signature.

1. FOR each voxel $v \in G$
2. $SubVolume = (W$ center at $v)$
3. Initialize a new feature histogram, $H[0...(N - \lceil \frac{N}{2} \rceil + 1)] = 0$
4. FOR each voxel $t \in SubVolume$
5. Calculate its corresponding BGSP Q_t with parameters R and N
6. Determine whether Q_t is a UGSP or not using Algorithm 1
7. IF Q_t is a UGSP
8. $S =$ Size of the largest connected component in Q_t
9. $PatternID = S - \lceil \frac{N}{2} \rceil$
10. $H[PatternID] = H[PatternID] + 1$
11. ELSE
12. $H[N - \lceil \frac{N}{2} \rceil + 1] = H[N - \lceil \frac{N}{2} \rceil + 1] + 1$
13. END IF
14. END FOR
15. Normalize $H[0...(N - \lceil \frac{N}{2} \rceil + 1)]$ such that $\sum_{i=0}^{N - \lceil \frac{N}{2} \rceil + 1} H[i] = 1$
16. $K(v) = H[0...(N - \lceil \frac{N}{2} \rceil + 1)]$
17. END FOR
18. Return K

between voxels. During the operation of finding the connected components in Algorithm 1, higher order voxel information is considered based on the first order information embedded in the label of each voxel. While LBP [8] only considers the first order information as it only compares the intensity differences between the neighboring voxels and the center voxel. Therefore, UGSP has higher order information layer than LBP.

In this paper, the type ID of a UGSP is determined based on the size of the largest connected component of the UGSP (e.g. $max(|C_1|, |C_2|)$, C_1 and C_2 are calculated in operation 7 in Algorithm 1, $|C_1|$ and $|C_2|$ respectively denote the sizes of C_1 and C_2). The procedure for calculating the UGSP feature signatures for each voxel of the input image is presented in Algorithm 2.

A more detail UGSP type classification can be achieved if we also consider the label type and shape of the largest connected component. However, it will make the UGSP feature histogram too sparse as there are too many possible combinations and some of the pattern's occurrence frequency are too small. Radius R in the Algorithm 2 affects the scale of interest. In this paper, the radius R is set by the best scale selection principle in [10].

3 MRF Modeling for Registration

In recent years, it is shown that the registration process can be robustly formulated as a MRF labeling problem [11]. In this paper, the UGSP feature is integrated with the MRF framework to drive the registration.

The general form of the MRF energy function can be defined by Equation 3:

$$\begin{aligned} E_f &= E_{data} + E_{smoothness} \\ &= \sum_{p \in \Omega} D_p(l_p) + \sum_{(p,q) \in N} V_{p,q}(l_p, l_q), \end{aligned} \tag{3}$$

where Ω is the set of voxels, N is the neighborhood system defined in Ω. In this paper, the 4-connected neighborhood system is used. $D_p(l_p)$ is the energy function associated with the data term. It penalizes the cost of assigning label l_p to voxel p. $V_{p,q}(l_p, l_q)$ is the energy function associated with the smoothness term, and it penalizes the cost of label discrepancy between two neighboring voxels.

The registration problem is transformed to the MRF labeling problem by quantizing the deformation space. A discrete set of labels $L \in \{l^1, l^2, ..., l^n\}$ is defined. Each label l^i $(1 \leq i \leq n)$ corresponds to a displacement vector $\boldsymbol{d_i}$. Assigning the voxel p with label l_p means moving p to a new position by the displacement vector $\boldsymbol{d_{l_p}}$. We follow the quantization step in [12], where each voxel can be moved from the original position bounded by a discretized window $\Psi = \{0, \pm s, \pm 2s, ..., \pm ws\}^d$ of dimension d.

The energy function $D_p(l_p)$ associated with the data term is defined based on the UGSP features as:

$$\begin{aligned} D_p(l_p) &= D_p(G_{template}(\boldsymbol{p}), G_{subject}(\boldsymbol{p} + \boldsymbol{d_{l_p}}) \\ &= D_p(K_{template}(\boldsymbol{p}), K_{subject}(\boldsymbol{p} + \boldsymbol{d_{l_p}})) \\ &= JSD(K_{template}(\boldsymbol{p}) || K_{subject}(\boldsymbol{p} + \boldsymbol{d_{l_p}})), \end{aligned} \tag{4}$$

where $G_{template}$ is the template image, $G_{subject}$ is the subject image, $K_{template}$ and $K_{subject}$ are the UGSP feature images of $G_{template}$ and $G_{subject}$ respectively at the current iteration calculated by Algorithm 2. $JSD(\cdot)$ is the Jensen-Shannon divergence measure.

The smoothness energy function $V_{p,q}(l_p, l_q)$ is defined as:

$$V_{p,q}(l_p, l_q) = \min(\lambda, |\boldsymbol{d_{l_p}} - \boldsymbol{d_{l_q}}|), \tag{5}$$

where λ is a constant represents the maximum penalty. It is the piece-wise truncated absolute distance.

The *alpha*-expansion algorithm [13] is applied to minimize the energy function defined in Equation 3 with the data term energy function and smoothness term energy function defined in Equations 4 and 5.

4 Experimental Results

The proposed method is evaluated on both the simulated and real 3D databases obtained from BrainWeb[1] and IBSR[2] respectively. In all experiments, the sub-volume window sizes W was set to $16 \times 16 \times 16$ in Algorithm 2, the number of neighboring samples N for each UGSP was 60. The 3D displacement window used in this paper was $\Psi = \{0, \pm 1, \pm 2, ..., \pm 12\}^3$. The maximum penalty parameter λ defined in Equation 5 was set to 20. The proposed method is also compared with three widely used approaches: FFD [3], Demons [4] and HAMMER [5]. In all the experiments, the control point spacing of FFD was set to 2.5mm. For Demons, the displacement field was smoothed by a unit variance Gaussian kernel. The elementary demon forces were computed by the optical flow equation [4].

4.1 Experiment on Simulated Data

Twenty image volumes from different subjects were obtained from BrainWeb. One of the image volumes was selected as the template image, and the others were used as the subject images. The resolution of each image was $256 \times 256 \times 181$ voxels. The segmentation results of each image for three classes of tissues: white matter (WM), gray matter (GM) and the cerebrospinal fluid (CSF) are also provided by the BrainWeb. The skull of each brain image was removed before registration by using the software Brain Suite version 2 obtained from USC [3] as it is a required step for HAMMER [5] to be compared in this paper.

The tissue overlap of GM, WM and CSF between the template and transformed subject images [14] was adopted as the evaluation function. The evaluation function is defined as $P = \frac{N(A \cap B)}{N(A \cup B)}$, where A and B denote the regions of a specific tissue in two images. The average values and standard deviations of P for GM, WM and CSF before registration, registration after using FFD [3], Demons [4], HAMMER [5] and the proposed method are listed in Table 1.

It is observed in Table 1 that the proposed method has the highest value of P among all the methods in this comparison. It reflects that the proposed method can give high registration accuracy.

4.2 Experiment on Real Data

The proposed method was also evaluated on the real datasets obtained from IBSR. Twenty skull-stripped image volumes with segmentation results were obtained from IBSR. The size of each volume was around $256 \times 256 \times 64$ voxels. The experimental settings were similar to the settings described in Section 4.1. Again, the tissue overlap evaluation function was used to measure the registration accuracy of different approaches. Table 2 lists the tissue overlap function value of P for various methods.

[1] http://www.bic.mni.mcgill.ca/brainweb/
[2] http://www.cma.mgh.harvard.edu/ibsr/
[3] http://brainsuite.usc.edu/

Table 1. The mean values of P and SDs of GM, WM and CSF with different methods on the simulated 3D database. BR denotes before registration.

Tissue	BR	FFD	Demons	HAMMER	UGSP
Gray	0.41923±0.07	0.75284±0.05	0.78522±0.04	0.80273±0.05	0.84452±0.04
White	0.48344±0.03	0.76409±0.06	0.78376±0.05	0.81296±0.04	0.83201±0.07
CSF	0.37025±0.06	0.72936±0.06	0.75784±0.04	0.76553±0.05	0.80146±0.05

Table 2. The mean values of P and SDs of GM, WM and CSF with different methods on the real 3D database. BR denotes before registration.

Tissue	BR	FFD	Demons	HAMMER	UGSP
Gray	0.54082±0.06	0.75193±0.05	0.77631±0.05	0.79063±0.06	0.83167±0.05
White	0.52147±0.05	0.76728±0.07	0.77382±0.06	0.80274±0.04	0.83639±0.04
CSF	0.33094±0.07	0.73425±0.05	0.76813±0.05	0.77153±0.05	0.80125±0.05

It is shown that the proposed method still has the largest value of P among all the methods in the comparison and it strongly implies the registration accuracy of the proposed method.

5 Conclusion

In this paper, a new feature based non-rigid image registration method is proposed. The proposed method is based on a new type of image feature, named uniform gradient spherical patterns (UGSP). UGSP encodes second order information to capture the geometric properties around each voxel. UGSP is rotation invariant and monotonic gray-level transformation invariant. The UGSP feature is integrated with the Markov random field (MRF) labeling framework to formulate the registration process. The proposed method is evaluated on both the simulated and real 3D datasets obtained from BrainWeb and IBSR respectively. It is also compared with other state-of-the-art registration methods. Experimental results show that the proposed method has the highest registration accuracies among all the compared methods on both the simulated and real 3D datasets.

References

1. Rohr, K.: Image registration based on thin-plate splines and local estimates of anisotropic landmark localization uncertainties. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 1174–1183. Springer, Heidelberg (1998)
2. Thompson, P., Toga, A.W.: A surface-based technique for warping three-dimensional images of the brain. TMI 15, 402–417 (1996)
3. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: Application to breast MR images. TMI 18, 712–721 (1999)

4. Thirion, J.P.: Image matching as a diffusion process: an analogy with maxwell's demons. MedIA 2, 243–260 (1998)
5. Shen, D., Davatzikos, C.: HAMMER: Hierarchical attribute matching mechanism for elastic registration. TMI 21, 1421–1439 (2002)
6. Tu, Z., Narr, K., Dollár, P., Dinov, I., Thompson, P., Toga, A.: Brain anatomical structure segmentation by hybrid discriminative/generative models. TMI 27, 495–508 (2007)
7. Dirk, L., Slagmolen, P., Maes, F., Vandermeulen, D., Suetens, P.: Nonrigid image registration using conditional mutual information. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 725–737. Springer, Heidelberg (2007)
8. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. PAMI 24, 971–987 (2002)
9. Yershova, A., LaValle, S.: Deterministic sampling methods for spheres and so(3). In: ICRA, pp. 3974–3980 (2004)
10. Wu, G., Qi, F., Shen, D.: Learning-based deformable registration of mr brain images. TMI 25, 1145–1157 (2006)
11. Glocker, B., Komodakis, N., Paragios, N., Tziritas, G., Navab, N.: Inter and intramodal deformable registration: Continuous deformations meet efficient optimal linear programming. In: IPMI, pp. 408–420 (2007)
12. Tang, T.W.H., Chung, A.C.S.: Non-rigid image registration using graph-cuts. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 916–924. Springer, Heidelberg (2007)
13. Yuri, B., Olga, V., Ramin, Z.: Fast approximate energy minimization via graph cuts. PAMI 23, 1222–1239 (2001)
14. Crum, W.R., Rueckert, D., Jenkinson, M., Kennedy, D., Smith, S.M.: A framework for detailed objective comparison of non-rigid registration algorithms in neuroimaging. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 679–686. Springer, Heidelberg (2004)

Methods for Tractography-Driven Surface Registration of Brain Structures

Aleksandar Petrović, Stephen M. Smith, Ricarda A. Menke, and Mark Jenkinson

Centre for Functional MRI of the Brain (FMRIB), University of Oxford
petrovic@fmrib.ox.ac.uk

Abstract. Registration of brain structures should bring anatomically equivalent areas into correspondence which is usually done using information from structural MRI modalities. Correspondence can be improved by using other image modalities that provide complementary data. In this paper we propose and evaluate two novel surface registration algorithms which improve within-surface correspondence in brain structures. Both approaches use a white-matter tract similarity function (derived from probabilistic tractography) to match areas of similar connectivity patterns. The two methods differ in the way the deformation field is calculated and in how the multi-scale registration framework is implemented. We validated both algorithms using artificial and real image examples, in both cases showing high registration consistency and the ability to find differences in thalamic sub-structures between Alzheimer's disease and control subjects. The results suggest differences in thalamic connectivity predominantly in the medial dorsal parts of the left thalamus.

1 Introduction

Different MRI modalities provide complementary data, informing us in various ways about brain anatomy and physiology. It is therefore natural to try to interpret them in an integrated fashion. This is beneficial for a variety of reasons; for example, matching structural landmarks (especially in the cortex) does not always imply matching functional regions [1]. Also, some structures appear featureless in a certain modality, with no particular structural landmarks for matching, whereas other modalities may provide clearer information (e.g. white matter changes in DTI data).

We believe that image registration can be improved if, in addition to matching T1-weighted image intensities, we also try to align brain regions according to their connectivity patterns (information from diffusion-weighted images). Most importantly, including connectivity information takes us closer to matching functional areas [2,3,4]. In this case, the description of every brain region becomes more specific, which should result in better matching accuracy across the population.

In this paper we propose two registration methods that integrate information available in T1-weighted images and probabilistic tractography. They use geometrical models of subcortical brain structures (e.g., human thalamus is used

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 705–712, 2009.

here) segmented using the FIRST/FSL software tool [5] to enhance anatomical correspondence between their surfaces. Anatomical alignment is based on inter-subject matching of areas of similar connectivity, which supposes that anatomically equivalent areas are connected to the same cortical and sub-cortical regions in different subjects. This is different from [4] where the cortical surface is deformed based on regional connectivity measures to cortical ROIs. We illustrate the effectiveness of our registration methods by applying them to an Alzheimer's disease (AD) data-set, specifically looking for thalamic connectivity alterations, which could have value in determining MRI biomarkers of disease progression.

2 Methods

The registration process tries to improve *within-surface* correspondence, which means that the shape of the extracted structures (represented by meshes output by FIRST) is unchanged by the surface matching. What changes is the correspondence between vertices of the two meshes, which represents anatomical correspondence between thalamic surfaces. For example, let every vertex in the thalamic mesh have a unique label (i.e., a number between 1 and N) and denote the set of vertices on the surface of the anterior dorsal thalamic nucleus as S_1 in one subject and S_2 in the other (in general, different sets of numbers). The purpose of the algorithm is to match vertices in S_1 to those in S_2 by sliding them around the surface and in that way re-establish correspondence of labels.

Both registration methods presented here use the same form for the similarity function which measures similarity in white-matter (WM) connectivity of certain brain regions to the rest of the brain. In our case, these brain regions are represented by different points of the mesh delineating the sub-cortical structure. Therefore, the first step in the processing pipeline is to segment matching structures (using FIRST) from T1-weighted images of both subjects (or more if group registration is performed). Meshes delineating segmented structures have a fixed number of vertices (e.g., the thalamus has 642 vertices) that, in our approach, become seeding points for whole-brain probabilistic tractography [6].

Next, we label probabilistic tractography output from vertex A as the "Connectivity Profile" of A: $\mathbf{CP}(A)$. Therefore, $\mathbf{CP}(A)$ is an image where every voxel encodes the number of white-matter tracts originating at coordinates of vertex A and passing through that voxel. The maximum value is the number of random tractography samples at the seed, which is 5000. In the next step, CPs of all vertices for both subjects' structures (human thalami) are transformed into standard MNI space where they can be compared.

After these steps the Correlation Matrix (CM) is formed, which encodes the similarity measure between CPs (reshaped into vectors). We define correlation of vectors $\mathbf{X}$ and $\mathbf{Y}$ as

$$Corr(\mathbf{X}, \mathbf{Y}) = \frac{\mathbf{X}^T\mathbf{Y}}{(\mathbf{X}^T\mathbf{X} + \mathbf{Y}^T\mathbf{Y})/2}, \tag{1}$$

so that the Correlation Matrix can be written as

$$CM(i,j) = Corr(\mathbf{CP}(\mathbf{x}_i), \mathbf{CP}(\mathbf{y}_j)), \;\; i,j \in [1,N], \tag{2}$$

where $\mathbf{x}_i$ is the ith vertex of a mesh delineating the thalamus in subject X ($\mathbf{y}_j$ for subject Y) and N is the number of vertices in each mesh. Values from the CM are used to derive the similarity function that drives the registration.

FIRST sub-cortical meshes are initially formed by deforming a spherical mesh. We use this equivalence in mesh topology to perform registration using spherical meshes. Registration starts by generating two identical regular spherical meshes (generated by subdividing an icosahedron) representing the input and the reference structure. The *reference mesh* (sphere) remains undeformed during registration whereas the *input mesh* undergoes tangential deformations to maximise the similarity function and better match the reference mesh. In this way, vertices from the input mesh move along the surface until they converge to a position where correlations between their CPs and the CPs of the corresponding vertices of the reference mesh are maximised. In the end, the deformation (of the input mesh) is propagated back from the sphere to the mesh of the structure in the native image space.

2.1 Registration Using Deformation Forces (*Method1*)

In this approach we make use of two deformation forces: the similarity force (SF) and the regularisation force (RF). The similarity force aims to maximise similarities between connectivity profiles, whereas the regularisation force aims to preserve the separation of the vertices (Fig. 1). A multi-scale registration framework is realised by grouping vertices into "patches". A "patch" represents a group of vertices that *move together*. Large patches are crucial at the initial stages of the registration when dealing with large-scale deformations e.g., global rotation. On the other hand, small patches (one vertex is the smallest) only act locally, affecting local mesh features. In total, optimisation is performed at 10 different scales (patch sizes).

The local similarity function (LSF) calculates the mean correlations between all vertices within a patch in the input mesh and the corresponding vertices of the

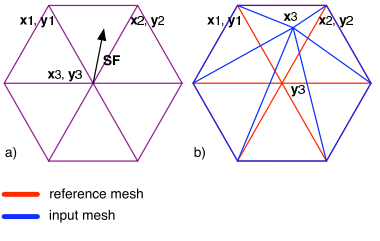


Fig. 1. The effect of the similarity force on an input mesh vertex ($\mathbf{x}_i$ – input mesh; $\mathbf{y}_i$ – reference mesh). The deformation of $\mathbf{x}_3$ due to the similarity force is shown. In a) reference and input meshes initially overlap. In b) the similarity force SF moves $\mathbf{x}_3$ away from $\mathbf{y}_3$ and towards $\mathbf{y}_2$ as $Corr(\mathbf{CP}(\mathbf{x}_3), \mathbf{CP}(\mathbf{y}_2)) > Corr(\mathbf{CP}(\mathbf{x}_3), \mathbf{CP}(\mathbf{y}_3))$. The regularisation forces would have the opposite effect, bringing $\mathbf{x}_3$ closer to $\mathbf{y}_3$.

reference mesh. The similarity force, **SF**, is defined as the spatial derivative of the LSF with respect to the parameters encoding the position of the patch centre in the input mesh. This force acts equally on all the vertices within a patch, inducing coherent movement. The regularisation force is defined as $\mathbf{RF}(\mathbf{x}_i, \nu_i) \propto \mathbf{m}_i - \mathbf{x}_i$ where ν_i is the set of first neighbours of $\mathbf{x}_i$ in X and $\mathbf{m}_i$ is the average position of the neighbours, given by $\mathbf{m}_i = \frac{1}{|\nu_i|} \sum_{k \in \nu_i} \mathbf{x}_k$ as in [7]. RF is calculated and applied separately to each vertex within a patch.

RF and SF are iteratively applied to patches until convergence. The patch sizes (number of vertices) we use are 642, 321, 100, 50, 40 30, 20, 10, 5, 3 and 1. At the smallest patch size, both forces act on one vertex in turn. During every iteration, each vertex becomes the centre of a patch once and is chosen at random. The number of iterations has been empirically optimised using artificial testing data as described below.

Displacements are determined from the forces by scaling parameters, λ and μ for RF and SF respectively, where the values are determined empirically. When a force is applied to a vertex, mesh-intersections and local topology breaches are prevented by checking for intersections and, when they occur, reducing the size of the movement until no intersections occur.

2.2 Registration Based on Spherical Wavelets (*Method2*)

In this registration method we represent the deformation field using a set of biorthogonal wavelet basis functions defined on a sphere [8,9]. The basis set is constructed of scaling functions at the coarsest scale and wavelet functions at finer scales. Both scaling (φ) and wavelet functions (ψ) are defined as $\varphi_{j,k}, \psi_{j,k} : \mathbb{S} \rightarrow \mathbb{R}$, where $\mathbb{S}$ is a unit sphere, j defines the scale of the function and k refers to the spatial index which describes where on the surface the function is centered. At a particular scale j, wavelet functions are combinations of scaling functions at scales j and $j+1$. A given function $f : \mathbb{S} \rightarrow \mathbb{R}$ can be expressed as a linear combination of the basis functions

$$f(\mathbf{x}) = \sum_k \lambda_{0,k} \varphi_{0,k}(\mathbf{x}) + \sum_{j \geq 0} \sum_m \gamma_{j,m} \psi_{j,m}(\mathbf{x}). \tag{3}$$

Scaling coefficients $\lambda_{0,k}$ represent the low-pass content of the signal f whereas coefficients $\gamma_{j,m}$ represent the localised band-pass content of the signal.

The idea of multi-scale registration using spherical wavelets relies on approximating the function f at different scales - coarse (low) scale encodes large deformations while higher scales determine fine features of the deformation field (for applications in shape analysis see, e.g., [10,11]). If a spherical mesh has N vertices, a total of N basis functions are created, composed of N_0 scaling functions (where N_0 is the initial number of vertices of the base mesh - e.g. icosahedron) and N_r wavelet functions. If each of these basis functions is evaluated in each of the N vertices and these data stacked into a matrix $\Phi_{N \times N}$, every finite energy scalar function F evaluated at N vertices can be transformed into a vector of basis coefficients Γ using Forward Wavelet Transform $\Gamma = \Phi^{-1} F$ and recovered using *Inverse Wavelet Transform*: $F = \Phi \Gamma$.

If F encodes displacements in Euler angles for every vertex of the input mesh, coefficients in Γ can be numerically solved for by maximising the similarity function between CPs of the input (X) and reference (Y) meshes. The similarity function (SFun) is calculated for all vertices of the reference mesh and can be written as

$$SFun(X,Y) = \mu \frac{1}{N} \sum_{l \in Y} \widetilde{CM}(\mathbf{x}_{l^*}, \mathbf{y}_l) + \nu E_m(F), \tag{4}$$

where $\widetilde{CM}(\mathbf{x}_{l^*}, \mathbf{y}_l)$ is the interpolated value of the CP correlation between $\mathbf{y}_l$ and the corresponding input mesh vertex $\mathbf{x}_{l^*}$, where this position directly depends on the deformation field, F. The interpolation of CM values is achieved using linear interpolation. E_m is the membrane energy of the deformation field, taking on the role of the regulariser, while μ and ν are scaling parameters found empirically. In order to penalise local topology breaches, parameter ν is increased (by 10%) when the deformation field F results in a mesh-intersection.

We use variable metric non-linear optimisation to estimate coefficients λ and ψ (Γ) [12]. This is performed in four steps (on four scales) with all coefficients set to 0 initially. We denote a set of coefficients at scale i by S_i, $i \in [0,3]$ (where $S_{i-1} \subset S_i$). The first stage of optimisation finds the coefficients at the coarsest scale, S_0, followed by stages for increasingly higher scales: S_1, S_2 and S_3.

3 Results

In order to test the algorithms and tune free parameters we ran experiments on a set of artificial examples. The artificial meshes are regular spheres with six differently coloured regions (Fig. 2). All vertices of the same colour have identical CPs across spheres/subjects, while vertices with different colours have completely uncorrelated CPs. In order to test the quality of registration we used all spheres from Fig. 2 and performed A to B, B to C and C to A registrations. If we denote the corresponding mappings by w_A, w_B and w_C: $A \xrightarrow{w_A} B \xrightarrow{w_B} C \xrightarrow{w_C} A$, then in an ideal case, $w_C \circ w_B \circ w_A = I$ where I is an identity mapping (we shall call this an $ABCA$ test). Therefore, the residual deformation field of the $ABCA$ test is a good measure of the registration consistency although it cannot be used to assess absolute accuracy. We performed the $ABCA$ test using Methods 1 and 2

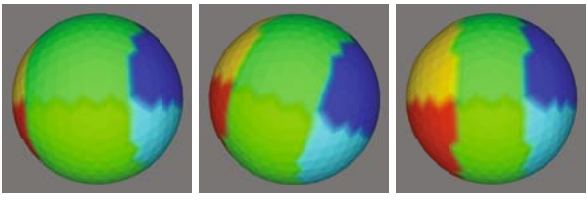


Fig. 2. A (left), B (middle) and C (right) represent three different artificial subjects/spheres. Sizes of the coloured regions and overall rotation vary between subjects.

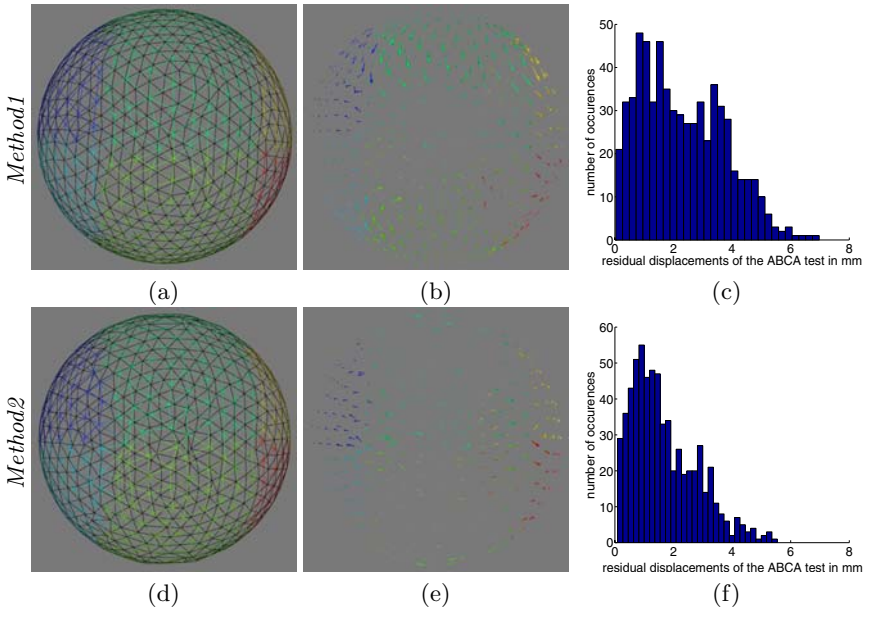
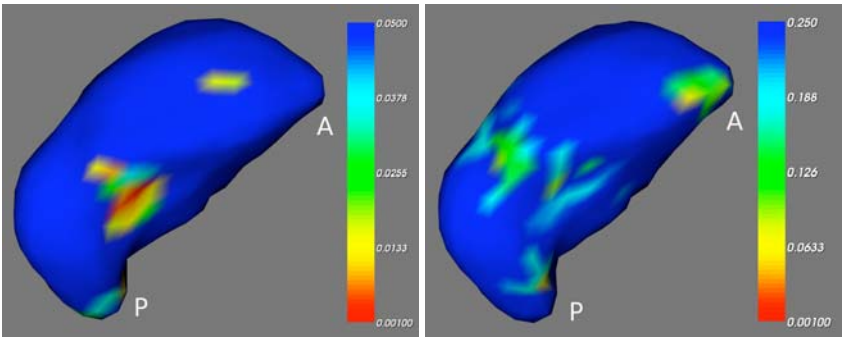


Fig. 3. *ABCA* test performed using Methods 1 and 2. a) shows mesh A after the $A \rightarrow B \rightarrow C \rightarrow A$ registration using *Method1*. In the ideal case, it should be a regularly tessellated sphere. b) shows displacement vectors between original A and A-registered-back-to-itself through B and C. c) shows a histogram of displacement vector norms from b) (mean displacement is approx. 2 mm). d), e) and f) show analogous results for *Method2*. All displacement values are smaller than the length of the mesh triangle side (≈7 mm). Mean (true) displacement values for each of the $A \rightarrow B$, $B \rightarrow C$ and $C \rightarrow A$ registration steps are 5 mm, 6 mm and 10 mm respectively for both methods.

and the results are shown in Fig. 3. For every registration step of the *ABCA* test, the final mean vertex-to-vertex correlations between the input and the reference meshes were approx. 0.98 (5% to 10% increase from the initial mean vertex-to-vertex correlation) indicating successful alignment/convergence for both registration methods. The *ABCA* test was also performed on thalami from three healthy controls. For *Method1*, the average vertex displacement from the registrations (A-B, B-C and C-A) was approx. 1.5 mm (5 mm max). The average residual displacements (A-B-C-A) were approx. 1 mm (5 mm max) and the average mesh edge length approx. 2 mm. As the ABCA test involves three registrations, the individual registration errors are likely to be less than the total ABCA residual error. For *Method2*, however, the average vertex displacements from the B-C and C-A registrations were similar to the average residual displacement, while the A-B registration found no displacements. This could be due to convergence problems with the variable metric approach when optimising fine-scale wavelet coefficients, possibly due to the highly non-linear similarity function. Conjugate gradient optimisation method yielded similar results.



(a) (b)

Fig. 4. Statistical analysis of residual within-surface displacements between control and AD groups for the left thalamus. Colourbars represent *uncorrected p-values*. a) Results for *Method1*. b) Results for *Method2*. A – anterior, P – posterior.

These methods were further tested on the thalami from a dataset comprising age- and gender-matched AD and control groups (12+12 subjects). An AD subject (the most typical of all 24 subjects [13]) was chosen as template and all others were registered to this using proposed methods. For every vertex, the Euclidean distance between its original position (given by FIRST vertex correspondence) and the new position after registration was compared between AD and control groups. For both methods two-tailed t-tests showed group differences in residual within-surface displacements predominantly in medial dorsal parts of the left thalamus (Fig. 4). Right thalamus showed negligible differences that considerably depended on the choice of the template subject. Spatial positions of changes found in the left thalamus did vary with the choice of template, but were always constrained to the medial dorsal part. Nevertheless, such dependencies demonstrate the potential benefits of construction and investigation of an average template of thalamic connectivity profiles.

4 Discussion

We presented a general framework for improving the *within-surface* correspondence between brain structures based on matching areas of similar structural connectivity patterns. We then introduced two methods implementing this idea. Both perform well on the artificial example suggesting high registration consistency and indicating that the chosen set of features is informative enough to drive the registration. *Method2*, however, was less sensitive when analysing real data, although it performed better in the artificial examples. We believe this is due to the difficult optimisation step at the highest scale (aligning fine details) where we found the variable metric optimisation approach was often unable to properly converge. This demonstrated the importance of adjusting the optimisation method, as well as the multi-scale approach, to the specific nature of this

registration problem. Nonetheless, patterns of differences in thalamic connectivity between AD and healthy controls found by both methods are consistent with histological evidence [14]. This indicates potential value for the method in determining MRI-based biomarkers of disease progression. Future work will attempt to apply this registration framework to the whole neo-cortical surface.

References

1. Amunts, K., Schleicher, A., Burgel, U., Mohlberg, H., Uylings, H.B., Zilles, K.: Broca's region revisited: cytoarchitecture and intersubject variability. J. Comp. Neurol. 412(2), 319–341 (1999)
2. Behrens, T., Jenkinson, M., Robson, M., Smith, S., Johansen-Berg, H.: A consistent relationship between local white matter architecture and functional specialisation in medial frontal cortex. NeuroImage 30(1), 220–227 (2006)
3. Johansen-Berg, H., Behrens, T., Sillery, E., Ciccarelli, O., Wheeler-Kingshott, C., Thompson, A., Smith, S., Matthews, P.: Functional-anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. Cerebral Cortex 15, 31–39 (2005)
4. Pascal, C., Jean-Francois, M.: Registration of cortical connectivity matrices. In: CVPRW 2006: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, Washington, DC, USA, p. 66. IEEE Computer Society Press, Los Alamitos (2006)
5. Woolrich, M., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.: Bayesian analysis of neuroimaging data in FSL. NeuroImage 45, S173–S186 (2009)
6. Behrens, T., Woolrich, M., Jenkinson, M., Johansen-Berg, H., Nunes, R., Clare, S., Matthews, P., Brady, J., Smith, S.: Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50(5), 1077–1088 (2003)
7. Smith, S.: Robust automated brain extraction. In: Sixth Int. Conf. on Functional Mapping of the Human Brain, vol. 625 (2000)
8. Schröder, P., Sweldens, W.: Spherical wavelets: efficiently representing functions on the sphere. In: SIGGRAPH, pp. 161–172 (1995)
9. Schröder, P., Sweldens, W.: Spherical wavelets: Texture processing. In: Hanrahan, P., Purgathofer, W. (eds.) Rendering Techniques, pp. 252–263. Springer, Heidelberg (1995)
10. Yu, P., Grant, P.E., Qi, Y., Han, X., Ségonne, F., Pienaar, R., Busa, E., Pacheco, J., Makris, N., Buckner, R.L., Golland, P., Fischl, B.: Cortical surface shape analysis based on spherical wavelets. IEEE Trans. Med. Imaging 26(4), 582–597 (2007)
11. Nain, D., Haker, S., Bobick, A.F., Tannenbaum, A.: Multiscale 3-d shape representation and segmentation using spherical wavelets. IEEE Trans. Med. Imaging 26(4), 598–618 (2007)
12. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: Numerical Recipes in C, 2nd edn. Cambridge University Press, Cambridge (1995)
13. Smith, S., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T., Mackay, C., Watkins, K., Ciccarelli, O., Cader, M., Matthews, P., Behrens, T.: Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. NeuroImage 31, 1487–1505 (2006)
14. Xuereb, J.H., Perry, R.H., Candy, J.M., Perry, E.K., Marshall, E., Bonham, J.R.: Nerve cell loss in the thalamus in alzheimer's disease and parkinson's disease. Brain 114(pt. 3), 1363–1379 (1991)

A Combined Surface And VOlumetric Registration (SAVOR) Framework to Study Cortical Biomarkers and Volumetric Imaging Data*

Eli Gibson, Ali R. Khan, and Mirza Faisal Beg

School of Engineering Science, Simon Fraser University
egibson@sfu.ca

Abstract. Constructing a one to one correspondence between whole brain MR image scans is a problem of critical importance in neuroimaging analyses. We present a framework to combine the strength of both surface-based and volumetric-based analyses for consistent, bijective data transfer between brain coordinate systems.

1 Introduction

Constructing a one to one correspondence between whole brain MR image scans is a problem of critical importance in neuroimaging analyses [1]. This underlies the ability to transform several types of data that are acquired or computed in the native brain's structural coordinates to a common central template brain space [2]. One type of data to transform is two-dimensional cortical surface-indexed measures such as cortical thickness, curvature, sulcal locations/depths or projected functional activation maps on the surface. A second type of data to be transformed are the 3D volumetric Cartesian grid-indexed measures such as structural images, or 3D functional activation maps from fMRI. Comprehensive neuroanatomical analysis of these two kinds of data necessitates that the distinct surface-indexed and Cartesian grid-indexed coordinate systems be consistently transformed bijectively between the brains being registered.

There now exist several volume to volume registration (VVR) methods that attempt to register brain images by minimizing the overlap of grayscale MR image intensities; these give good results for most internal areas of the brain but cannot handle the thin and highly convoluted nature of the cerebral cortex. On the other hand, surface to surface registration (SSR) methods are able to preserve cortical topology and still give good registrations, however, either these lead to large distortions in the cortical mesh, or require quality controlled point or curve landmarks, the variability in natural occurrence or the uncertainty injected in

* We are grateful to NSERC, CHRP, Pacific Alzheimer Research Foundation and the Michael Smith Foundation for Health Research for providing funding towards this research.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 713–720, 2009.

the labeling process of which also influences heavily the quality of this registration. More generally, these SSR methods give cortical surface correspondence, but these are not easily extendable to construct a volumetric matching between the spaces. Hence, the important bridge that should exist between the two methods is not yet available. Instead, researchers have tended to use one or the other type of data and the corresponding registration, and have made comparisons to justify which one is suitable in what conditions [3], [4] and for surface-indexed measurements, generally conclude that surface based registrations offer better alignment than volumetric registration methods [5]. Some hybrid methods leverage VVR to bootstrap SSR [6]; others use SSR to initialize VVR [7,8] improving cortical alignment; however, these methods still use either VVR or SSR alone as the final registration.

Instead of choosing between grid-indexed or surface-indexed data while studying neuroanatomy, a better solution is a consistent combination of both of these methods to reach a comprehensive neuroimaging analyses framework. What is needed is a method that performs accurate whole brain matching such that volumetric and surface data can be consistently mapped to a given template space. The construction of such a framework, that yields consistent mapping between cortical surfaces and volumetric domains, is the focus of this paper.

2 Proposed Combined Surface And VOlumetric Registration (SAVOR) Framework

The neuroanatomical volumetric coordinate system Ω_i in the brain scan I_i of the i^{th} brain consists of the Cartesian grid-based coordinates indexing volumetric subdomains $V_i \subset \Omega$, and additionally, imbedded in Ω_i are the 2-dimensional surface manifolds S_i representing the whole brain cortical surface. Our method for constructing a whole brain and surface bijective and accurate correspondence

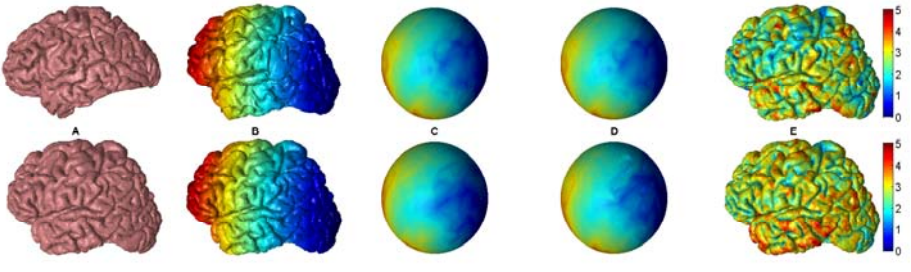


Fig. 1. Framework Pipeline. (A) Two pial surfaces (B) Upper surface registered to lower surface via image registration, with identity coordinate functions overlaid (C) surfaces mapped to the spherical domain via FreeSurfer mapping with coordinate functions overlaid (D) surfaces on spherical domain with function on lower surface mapped to that on upper surface (E) Cortical thickness values for upper and lower brains displayed on the domain of the lower brain.

relies on accurate VVR, followed by a surface constrained optimal approximation of the VVR using SSR. This process is outlined in Figure 1.

Whole Brain Smooth and Invertible VVR. The first step (Figure 1(B)) is to use an accurate whole brain VVR method construct a dense volumetric correspondence $\varphi_{V(i,j)} : \Omega_i \rightarrow \Omega_j$. This mapping should accommodate the variability of subcortical structures, as well as do a reasonably good job, although not necessarily perfect, of aligning cortical surfaces. This mapping will be used directly for transforming grid-index data. Additionally, this mapping is used as a reference registration that the following SSR will approximate.

Smooth and Invertible SSR. The second step is to construct invertible mapping $\varphi_{S(i,j)}$ between the two surfaces $S_i \in \Omega_i$ and $S_j \in \Omega_j$, proceeding from this dense volumetric mapping $\varphi_{V(i,j)}$ using VVR in step 1. Using the $\varphi_{V(i,j)}$ obtained, get a transformed surface $S_{i'} = \varphi_{V(i,j)}(S_i) = \{\varphi_{V(i,j)}(x), \forall x \in S_i\}$; this now lies in the domain Ω_j. The transformed surface $S_{i'}$ is expected to be close to the target surface S_j but not exactly matched. To achieve an accurate surface to surface correspondence, we need to construct a second mapping, $\varphi_{Si',j}$, between $S_{i'}$ and S_j to be used to ultimately construct the mapping from S_i to S_j. This mapping should approximate $\varphi_{V(i,j)}$ as closely as possible, under the constraint of precisely, aligning the surface domains. We propose the following three steps.

1. Define suitable functions indexed on the surfaces $S_{i'}$ and S_j that should be registered (Figure 1(B)),
2. define a smooth and invertible mapping of each surface to a common intermediate surface where these registrable functions defined in (a) are then transferred (Figure 1(C)),
3. compute a registration of these transformed functions on the common intermediate surface domain (Figure 1(D)).

The functions we choose to construct for registration are the coordinate identity map on $S_{i'}$ and S_j, such that $C(S_{i'}(x)) = x$ and $C(S_j(x)) = x$. These are then mapped to a common surface domain using a surface homeomorphic mapping $\varphi_{S(i',0)}$ and $\varphi_{S(j,0)}$ for each surface $S_{i'}$ and S_j to a common space S_0 along with the coordinate functions from both surfaces. In this common domain, perform a multi-dimensional registration of the mapped coordinate functions from the two surfaces, producing a mapping $\varphi_{S(0)}$. This mapping between surface coordinate identify maps transferred to a common domain in effect minimizes the distance between corresponding original surface points. Once these coordinate map functions are matched on the common domain via $\varphi_{S(0)}$, then an invertible mapping between the surfaces $S_{i'}$ and S_j is created by composition. By further composition with mapping between Ω_i and Ω_j, the final invertible map that preserves topology between the surfaces S_i and S_j is given to be:

$$\varphi_{S(i,j)} = \varphi^{-1}_{S(j,0)} \circ \varphi_{S(0)} \circ \varphi_{S(i',0)} \circ \varphi_{V(i,j)} \tag{1}$$

Interpolation and Transfer of Coordinate-indexed Data. The third step is interpolation of surface-indexed data to a common domain; using the surface mapping $\varphi_{S(i,j)}$ to transform the surface-index data. In the general case, where the surface-indexed functions are continuously defined on the surface domains, and all the mappings are continuous mappings between domains, there is no need for interpolation. However, implementations of this proposed framework may use sparse representations of the domain (as in the triangulated mesh-based surfaces in the illustrative example below). In this case, interpolation is necessary, as the point $x' \in S_j$ corresponding to point $x \in S_i$ will in general not lie on a point where the data to transform, F_{S_j}, is defined. This final interpolation step completes the transfer of surface-indexed functions between surfaces.

3 Our Implementation of the Proposed Framework

In implementing the framework, several algorithmic components must be selected: the VVR method for volumetric image registration, the homeomorphic mapping to bring surfaces into to a common domain, the method of registration of coordinate functions on this common domain and finally, the method of interpolation. Our choice of these components and the selection rationale follow.

Data. T1-weighted MR images (N=8 subjects), each scanned at 2 time points were taken from the ADRC dataset [9]. Subjects having CDR 0 score at baseline and CDR 0.5 at followup were chosen. Three types of comparison were implemented: 1) transforming cortical surface-index data from each subject to a common template (cross-sectional); 2) transforming data from the followup scan for each subject to compare to the baseline data (longitudinal); and finally a combined approach to compare these longitudinal results across subjects in a common template. For each brain, the cortical pial surface, corresponding to Ω_{S_i}, was extracted using FreeSurfer [10]. Biomarkers defined on these surfaces include curvature measures generated by FreeSurfer, and our in-house implementation of cortical thickness measurements measured using Laplacian streamlines [11].

Image Registration. Accuracy of the VVR registration is important to the quality of the final mapping; thus a high-dimensional nonlinear VVR method was chosen. Specifically, VVR registration between MR images I_i and I_j was carried out using a multi-structure extension of a high dimensional nonrigid diffeomorphic (smooth and invertible) transformation [12] that incorporated channels for MR image intensity, subcortical binary segmentation and volumetric cortical segmentations to concurrently guide the registration. The simultaneous usage of the automated segmentations as separate cost terms allows the overall MR image matching to better avoid local minima, while providing flexibility in setting weights for different channels to emphasize certain properties, such as emphasizing smaller structures, or deemphasizing less reliable channels. The use of 34 cortical parcellations, as listed in [13] and computed using FreeSurfer, allows for higher accuracy in the cortical mantle. Parcellations were voxelized, then smoothed to eliminate the sharp boundaries, and used in guiding the

volumetric registration. This multi-structure framework is observed to transform the surfaces to within 1.5 mm of template surface in cross-sectional registration.

Homeomorphic Surface to Common Domain Mapping. To apply the second step for SSR, we map the surfaces to a common spherical domain. The specific mapping to a spherical domain is drawn from an intermediate step in the FreeSurfer segmentation, and is described in detail in [14]. In brief, each cortical surface is first inflated to a smooth surface and then projected on to a sphere. The surface is then evolved to minimize metric distortion, including spherical folds introduced by the projection.

Registration of Functions on Common Domain. The registration of the coordinate functions on the spherical domain is performed using Spherical Demons registration [15]. The Spherical Demons algorithm uses representations of mappings, and regularizations that are particularly suited to fast computation on the spherical domain, and know to perform well on cortical surfaces.

Interpolation. Each cortical surface was represented by a triangular mesh, and surface-indexed function was given on the vertices. Additionally, the homeomorphic mappings to the common domain are defined only at the vertices, extended via linear interpolation along the entire face, and for interpolating function value from neighboring vertices. This linearity allows us to compute the interpolation weights from a discretization of the spherical common domain.

4 Results

Using our multi-structure VVR registration, the Dice coefficients found for subcortical structure mappings are shown in Figure 3(b). For the subsequent SSR registration of the cortical surface manifolds between the baseline scans to a common template image(I_T)/surface(S_T) and from baseline scans to the followup scan, we transformed all baseline surfaces S_i and corresponding baseline FS labels F_i to the template domain. We compute, for each vertex v on the template surface, the surface registration error map (SREM) by comparing the transformed labels $F_{i,T}$ to the template label F_T via $\delta_{(F_T=F_{i,T})}$ giving value 1 for a mismatch, and 0 otherwise. Across N subjects, we define the probability of registration error to be the average SREM or ASREM via $P(error(v) = 0) = (1/N)\sum_{i=1}^{N} \delta_{(F_T=F_{i,T})}$. The ASREM, shown in Figure 2 (a) shows the average registration errors across our 8 subjects, and panel (b) shows the individual SREMs in matching to template space and (c) shows the SREMs for baseline to longitudinal matching. Using these registrations, we transformed the cortical curvature and mantle thickness functions longitudinally to baseline, and further, cross-sectionally to template. Longitudinal cortical thickness change, normalized for inter-scan time for each subject, is shown in Figure 4(a), and mapped further to a common template, as in Figure 4(b). Pearson's linear correlation, comparing the overlap of values transformed from followup to baseline, was done using our framework and using solely FreeSurfer's cortical surface registration [14], is shown in Figure 3(a). Dice metrics for alignment of cortical parcellations are given in Figure 3(b).

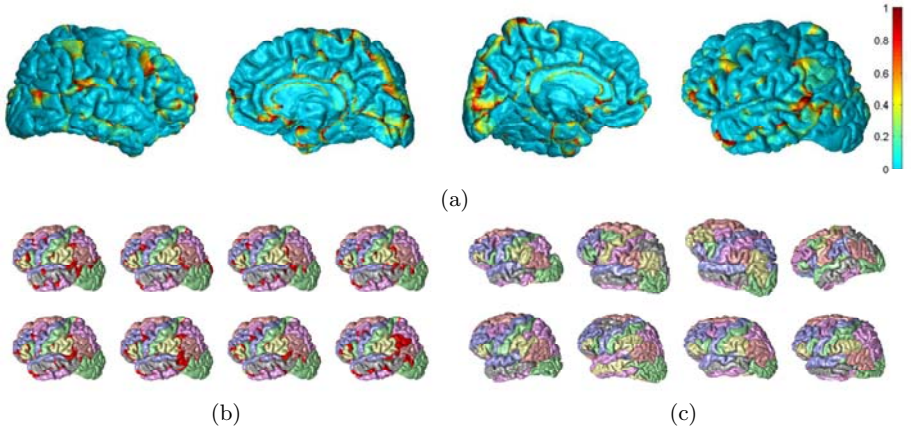
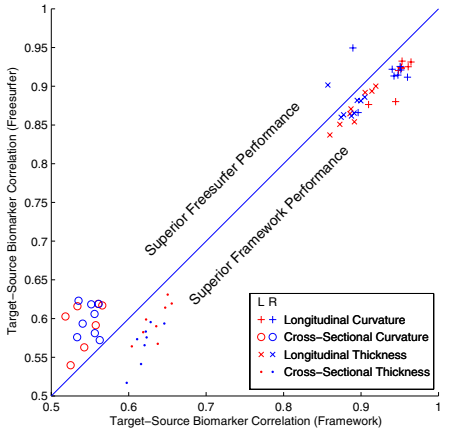


Fig. 2. The average surface-registration error map (ASREM), shown as a color map (a) on the template surface, quantifies the variability in label overlaps after registration, capturing the variability in folding patterns accuracy in the assignment of labels and errors in registration. (b) shows the individual cross-sectional registration error (red: labels were not matched), and (c) shows the longitudinal registration errors.



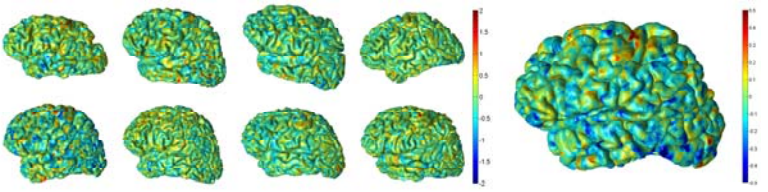
(a)

Gyri	Cross-sect.	Long.
Precentral	.94(.01)	.99(.01)
Postcentral	.93(.02)	.99(.01)
Sup. Frontal	.94(.03)	.99(.01)
Sup. Temporal	.91(.03)	.99(.01)
Inf. Temporal	.86(.04)	.99(.01)
Sup. Parietal	.92(.02)	.99(.01)
Inf. Parietal	.87(.08)	.99(.01)

Structure	Cross-sect.	Long.
White Matter	.66(.03)	.80(.12)
Ventricle	.85(.04)	.91(.02)
Thalamus	.85(.04)	.92(.02)
Caudate	.71(.06)	.83(.07)
Putamen	.78(.04)	.90(.02)
Hipocampus	.73(.06)	.85(.05)
Amygdala	.73(.07)	.86(.05)

(b)

Fig. 3. (a) Comparison of our framework vs. FreeSurfer cortical registration. Points below the line correspond to mappings where our framework had higher correlation between source and mapped target data. (b top) Mean Dice metrics for selected surface parcellations. (b bottom) Mean Dice metrics for selected subcortical structures.



(a) (b)

Fig. 4. (a)Longitudinal change in cortical thickness overlaid on the baseline surface for left hemisphere of all subjects. (b)Mean longitudinal change in cortical thickness mapped on to the common template. These thickness maps have not been smoothed.

5 Discussion and Conclusion

While direct surface registration has many proponents, there are limitations to the types of data that can be expressed on surfaces, and to the quality of registration achievable by directly mapping two surfaces. Furthermore, this step necessarily assumes that there is no error in the definition of the surface. Hence, the 'correct' registration may lie outside of the space of considered registrations in direct SSR. In the SAVOR framework, the correct registration would lie within the space of considered registrations, and errors in surface definition would be accounted for by the manifold registration on the common domain.

Our suggestion to use an arbitrary homeomorphic mapping to a common domain as an intermediate stage for mapping surfaces needs some justification. In the general sense, the mapping may be arbitrary; however, this requires that the manifold registration must also be capable of producing arbitrary mappings. In practice, manifold registration algorithms use regularization techniques that limit the space of achievable manifold registrations, limiting the choice of acceptable homeomorphic mappings. The Spherical Demons algorithm was used with the FreeSurfer spherical mapping based on evidence that this registration was powerful enough to align cortical surfaces with this mapping [15].

Our demonstrated implementation performed better at overlaying longitudinal cortical thickness and data mapping. For longitudinal and cross-sectional mapping of thickness and curvature, our implementation performed better except for cross-sectional curvature mapping. This may be attributed to the underlying mechanism of the FreeSurfer data mapping which includes a minimization of intersubject curvature in its cortical registration.

In conclusion, we have introduced a comprehensive framework for bijectively transporting both volume-indexed and surface-indexed data using a powerful combination of image-based and surface-based registration (SAVOR). We demonstrate the use of this framework for assessing general trends in longitudinal studies, and show results from one such assessment. This work simplifies the use of cortical surface biomarkers for proponents of image-based registration; enables combined surface and volume based data-mapping; and provides a new means for evaluating image-based registrations relative to surface-based registrations. However, the framework relies on a set of subcomponents, and the full potential

of this framework can be realized with optimized combination of each of the subcomponents.

References

1. VanEssen, D.C., Drury, H., Joshi, S.C., Miller, M.I.: Functional and structural mapping of human cerebral cortex: Solutions are in the surfaces. Proc. Nat. Acad. Sci. 95, 788–795 (1998)
2. Mangin, F., Riviére, J., Cachia, D., Duchesnay, A., Cointepas, E., Papadopoulos-orfanos, Y., Scifo, D., Ochiai, P., Régis, T., Erm, J.I.: A framework to study the cortical folding patterns. Neuroimage 23, 129–138 (2004)
3. Desai, R., Liebenthal, E., Possing, E.T., Waldron, E., Binder, J.R.: Volumetric vs. surface-based alignment for localization of auditory cortex activation. Neuroimage 26(4), 1019–1029 (2005)
4. Anticevic, A., Dierker, D., Gillespie, S., Repovs, G., Csernansky, J., VanEssen, D., Barch, D.: Comparing surface-based and volume-based analyses of functional neuroimaging data in patients with schizophrenia. Neuroimage 41(3), 835–848 (2008)
5. Fischl, B., Sereno, M.I., Tootell, R.B.H., Dale, A.M.: High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping 8(4), 272–284 (1999)
6. Liu, T., Shen, D., Davatzikos, C.: Deformable registration of cortical structures via hybrid volumetric and surface warping. Neuroimage 22(4), 1790–1801 (2004)
7. Eldeib, A., Yamany, S., Farag, A.: Volume registration by surface point signature and mutual information maximization with applications in intra-operative MRI surgeries. In: Proc. Int. Conf. Image Processing, September 2000, vol. 1, pp. 200–203 (2000)
8. Postelnicu, G., Zollei, L., Fischl, B.: Combined volumetric and surface registration. IEEE Trans. Med. Imaging 28(4), 508–522 (2009)
9. Alzheimer's Disease Research Center, Washington University, S.L.: Healthy aging & senile dementia program project (2004-2008)
10. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron. 33(3), 341–355 (2002)
11. Jones, S.E., Buchbinder, B.R., Aharon, I.: Three-dimensional mapping of cortical thickness using laplace's equation. Human Brain Mapping 11(1), 12–32 (2000)
12. Beg, M.F., Miller, M., Trouve, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision 61(2), 139–157 (2005)
13. Desikan, R., Ségonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, P., Hyman, B., Albert, M., Killiany, R.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31(3), 968–980 (2006)
14. Fischl, B., Sereno, M.I., Dale, A.M.: Cortical surface-based analysis - ii: Inflation, flattening, and a surface-based coordinate system. Neuroimage 9(2), 195–207 (1999)
15. Yeo, B.T.T., Sabuncu, M.R., Vercauteren, T., Ayache, N., Fischl, B., Golland, P.: Spherical demons: Fast surface registration. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 745–753. Springer, Heidelberg (2008)

Fast Tensor Image Morphing for Elastic Registration

Pew-Thian Yap[1,3], Guorong Wu[1,3], Hongtu Zhu[2,3], Weili Lin[1,3], and Dinggang Shen[1,3]

[1] Department of Radiology, [2] Department of Biostatistics and [3] BRIC
University of North Carolina at Chapel Hill, NC

Abstract. We propose a novel algorithm, called *Fast Tensor Image Morphing for Elastic Registration* or F-TIMER. F-TIMER leverages multiscale tensor regional distributions and local boundaries for hierarchically driving deformable matching of tensor image volumes. Registration is achieved by aligning a set of automatically determined structural landmarks, via solving a soft correspondence problem. Based on the estimated correspondences, thin-plate splines are employed to generate a smooth, topology preserving, and dense transformation, and to avoid arbitrary mapping of non-landmark voxels. To mitigate the problem of local minima, which is common in the estimation of high dimensional transformations, we employ a hierarchical strategy where a small subset of voxels with more distinctive attribute vectors are first deployed as landmarks to estimate a relatively robust low-degrees-of-freedom transformation. As the registration progresses, an increasing number of voxels are permitted to participate in refining the correspondence matching. A scheme as such allows less conservative progression of the correspondence matching towards the optimal solution, and hence results in a faster matching speed. Results indicate that better accuracy can be achieved by F-TIMER, compared with other deformable registration algorithms [1, 2], with significantly reduced computation time cost of 4–14 folds.

1 Introduction

Diffusion tensor imaging (DTI) is capable of measuring water diffusion in vivo non-invasively, and has been widely employed to delineate potential neurological disease related white matter abnormalities. To this end, accurate registration of diffusion tensor images across different subjects is a critical prerequisite for detailed voxel-by-voxel statistical analysis. Spatial normalization of diffusion tensor images is, however, challenging both technically and computationally given that tensor data representation is inherently high dimensional, and the anisotropic nature of cellular water diffusivity calls for proper reorientation of the tensors on top of their spatial alignment, which adds another dimension of difficulty to the problem.

Conventional methods generally extract tensor scalar features from each tensor individually, and by constructing scalar maps, regional integration and other operations such as edge detection can be performed to extract the final features

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 721–729, 2009.

for correspondence matching. These approaches, however, do not exploit full tensor information, and are limited in the sense that the regional and edge information extracted might not reflect the true underlying tensor structure. While it is possible to use existing scalar image registration algorithms to establish correspondence matching, it is natural to expect that the inclusion of additional information gained from considering the full diffusion tensor will lead to more accurate matching, and hence more robust registration.

In this paper, we propose a fast and accurate diffusion tensor image registration algorithm, called F-TIMER. F-TIMER, in addition to scalar map based features, gathers directly from the tensors regional statistical information and boundary edge information in a multiscale fashion. All the collected information is then grouped, for each voxel, into an *attribute vector* [3], serving as its structural signature. Based on the attribute vectors, salient points, signifying important anatomical structures, are automatically selected, and are utilized as landmarks for correspondence matching. For robust correspondence matching and to avoid being bias towards the template or the subject, the deformation is estimated by jointly considering the forward and reverse transformations [4]. Upon establishing landmark correspondences, thin-plate splines (TPS) [5] are employed to generate a smooth, topology preserving, and dense transformation, and to avoid arbitrary mapping of non-landmark voxels. The whole framework can be summarized as an *Attributive Symmetric Soft Assignment Problem* (ASSAP), which is an extension of the original soft assignment problem [6] by taking into account attribute vector labelled landmarks [3] and also symmetric transformation [4]. To obviate local minima, which are prone to happen, and often inevitable in estimating a transformation with very high degrees of freedom, we adopt a hierarchical strategy in optimizing the ASSAP related energy function. We progressively build up our optimization solution starting with landmarks resulting from a small subset of the voxels, which exhibit more distinctive attribute vectors — equivalent to optimizing an approximate lower-degrees-of-freedom version of the energy function — and include more voxels to refine the solution as registration progresses. Finally, reorientation of the tensors is performed using the algorithm presented in [7]. Such scheme, as can be validated from the experiments, results in more robust and accurate matching in a lesser amount of time.

2 Methods

2.1 Attribute Vectors

In medical imaging, it is important to build deformable anatomical models that take into account the underlying anatomy, and not simply the similarity of image intensities. To this end, a multiscale attribute vector is attached to each voxel, reflecting its underlying anatomical structure in a local scale, and also its relationship to more distant voxels in a more global scale. A rich enough attribute vector can potentially differentiate different parts of the anatomy that would otherwise look similar. F-TIMER characterizes a voxel using three different types of features: 1) Regional features $\mathbf{a}_r^{\mathrm{Reg}}(\mathbf{x})$ (means and variances),

2) Edge features $\mathbf{a}_r^{\mathrm{Edg}}(\mathbf{x})$ (tensor edges and FA map edges), and 3) Geometrical features $\mathbf{a}_r^{\mathrm{Geo}}(\mathbf{x})$ (FA values and principal diffusivities). These features are computed in three different scales (Fine, Middle, Coarse) and are grouped, for each voxel $\mathbf{x}$, into an attribute vector: $\mathbf{a}(\mathbf{x}) = \left[\mathbf{a}_{\mathrm{Fine}}(\mathbf{x}), \mathbf{a}_{\mathrm{Middle}}(\mathbf{x}), \mathbf{a}_{\mathrm{Coarse}}(\mathbf{x})\right]$, where, for each scale $r \in \{\mathrm{Fine}, \mathrm{Middle}, \mathrm{Coarse}\}$, there are three types of features: $\mathbf{a}_r(\mathbf{x}) = \left[\mathbf{a}_r^{\mathrm{Reg}}(\mathbf{x}), \mathbf{a}_r^{\mathrm{Edg}}(\mathbf{x}), \mathbf{a}_r^{\mathrm{Geo}}(\mathbf{x})\right]$. In the following, we describe and define $\mathbf{a}_r^{\mathrm{Reg}}(\mathbf{x})$, $\mathbf{a}_r^{\mathrm{Edg}}(\mathbf{x})$, $\mathbf{a}_r^{\mathrm{Geo}}(\mathbf{x})$, and also measures for gauging attribute vector similarity.

Regional Features. Utilizing Log-Euclidean metrics, we can define the regional mean in the neighborhood $\mathcal{N}(\mathbf{x})$ of voxel $\mathbf{x}$ as: $\mathbf{M}(\mathbf{x}) = \exp\left[\frac{\sum_{\mathbf{z}\in\mathcal{N}(\mathbf{x})} \log(\mathbf{D}(\mathbf{z}))}{|\mathcal{N}(\mathbf{x})|}\right]$, where $\mathbf{D}(\mathbf{x})$ is the tensor pertaining to voxel $\mathbf{x}$. From the mean, the principal diffusivities, i.e., the eigenvalues, can be computed: $\lambda_1^{(\mathbf{M})}(\mathbf{x}) \geq \lambda_2^{(\mathbf{M})}(\mathbf{x}) \geq \lambda_3^{(\mathbf{M})}(\mathbf{x})$, where $\lambda_k^{(\mathbf{M})}(\mathbf{x})$ represents the k-th largest eigenvalue of matrix $\mathbf{M}(\mathbf{x})$. Similarly, we can define the regional variance as: $\underline{\mathbf{V}}(\mathbf{x}) = \left[\sum_{\mathbf{z}\in\mathcal{N}(\mathbf{x})} \frac{[\log(\mathbf{D}(\mathbf{z})) - \log(\mathbf{M}(\mathbf{x}))]^2}{|\mathcal{N}(\mathbf{x})|}\right]$, and the principal variabilities as: $\lambda_1^{(\mathbf{V})}(\mathbf{x}) \geq \lambda_2^{(\mathbf{V})}(\mathbf{x}) \geq \lambda_3^{(\mathbf{V})}(\mathbf{x})$. These eigenvalues are scaled according to the following equation to yield their mean normalized values: $\tilde{\lambda}_i^{(\mathbf{V})}(\mathbf{x}) = \lambda_i^{(\mathbf{V})}(\mathbf{x}) / \sum_{k=1}^{3} \lambda_k^{(\mathbf{M}^2)}(\mathbf{x})$ where we have used $\underline{\mathbf{M}}^2$ instead of $\underline{\mathbf{M}}$ to match the dynamic range of $\underline{\mathbf{V}}$. The notation $\underline{[\cdot]}$ indicates that the quantity in the bracket is in the log space. The regional features are computed for each scale r, and are grouped into: $\mathbf{a}_r^{\mathrm{Reg}}(\mathbf{x}) = \left[\lambda_{1,r}^{(\mathbf{M})}(\mathbf{x}), \lambda_{2,r}^{(\mathbf{M})}(\mathbf{x}), \lambda_{3,r}^{(\mathbf{M})}(\mathbf{x}), \tilde{\lambda}_{1,r}^{(\mathbf{V})}(\mathbf{x}), \tilde{\lambda}_{2,r}^{(\mathbf{V})}(\mathbf{x}), \tilde{\lambda}_{3,r}^{(\mathbf{V})}(\mathbf{x})\right]$.

Edge Features. To better extract tissue boundaries, the Canny edge detector is extended to cater for diffusion tensor images. For fast edge detection, 3D Gaussian-based image filtering is implemented using three subsequent steps of one-dimensional (1D) Gaussian filtering along the anterior-posterior, superior-inferior and left-right directions, which is then followed by gradient maps computation. Using these steps, edge detection can be accomplished rapidly and robustly. Note that tensor edge detection is performed in the logarithmic space [8]. For each voxel in the volume, a gradient $\mathbf{H}^{\mathrm{Tensor}}(\mathbf{x})$ can be computed, and from which, after non-maximum suppression, a final edge magnitude $H^{\mathrm{Tensor}}(\mathbf{x})$ can be obtained. Edges from tensors and edges from FA map are complementary to each other and, by using both, potentially all major kinds of tissue boundaries, that is, those formed between white matter (WM), gray matter (GM) and cerebro-spinal fluid (CSF), can be detected and aligned in the registration. We denote the edge magnitude returned by the FA map detection at point $\mathbf{x}$ as $H^{\mathrm{FA}}(\mathbf{x})$. For scale r, the edge features are grouped into: $\mathbf{a}_r^{\mathrm{Edg}}(\mathbf{x}) = \left[H_r^{\mathrm{Tensor}}(\mathbf{x}), H_r^{\mathrm{FA}}(\mathbf{x})\right]$.

Geometrical Features. The rest of the features used in F-TIMER includes the fractional anisotrophy value $\mathrm{FA}(\mathbf{x})$ and the principal diffusivities $\lambda_1^{(\mathbf{D})}(\mathbf{x}) \geq \lambda_2^{(\mathbf{D})}(\mathbf{x}) \geq \lambda_3^{(\mathbf{D})}(\mathbf{x})$, which characterize the geometrical shape of the tensor ellipsoid. For scale r, the geometrical features are grouped into: $\mathbf{a}_r^{\mathrm{Geo}}(\mathbf{x}) = \left[\mathrm{FA}_r(\mathbf{x}), \lambda_{1,r}^{(\mathbf{D})}(\mathbf{x}), \lambda_{2,r}^{(\mathbf{D})}(\mathbf{x}), \lambda_{3,r}^{(\mathbf{D})}(\mathbf{x})\right]$.

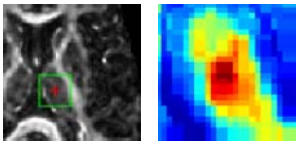


Fig. 1. Distintiveness of attribute vector. The similarity map in the green box is magnified for a closer inspection. Dark red indicates high similar and dark blue otherwise.

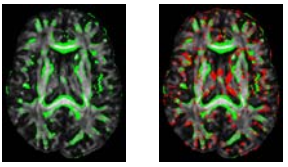


Fig. 2. Landmarks superimposed on the FA image. Shown on the left are the initial landmarks and on the right, in red, are the additional landmarks when the registration progresses to a later stage.

Similarity Measures. Prior to computing the attribute vector similarity, we normalize the elements of the attribute vectors to have a range of $[0, 1]$. For a template voxel $\mathbf{x}_t$ with attribute vector $\mathbf{a}_{\mathrm{T}}(\mathbf{x}_t)$ and a subject voxel $\mathbf{x}_s$ with attribute vectors $\mathbf{a}_{\mathrm{S}}(\mathbf{x}_s)$, their voxel similarity is defined as: $m\big(\mathbf{a}_{\mathrm{T}}(\mathbf{x}_t), \mathbf{a}_{\mathrm{S}}(\mathbf{x}_s)\big) = \mathit{\Pi}_i(1 - |a_{\mathrm{T},i}(\mathbf{x}_t) - a_{\mathrm{S},i}(\mathbf{x}_s)|)$, where $a_{\mathrm{T},i}(\mathbf{x})$ and $a_{\mathrm{S},i}(\mathbf{x})$ are the i-th elements of $\mathbf{a}_{\mathrm{T}}(\mathbf{x})$ and $\mathbf{a}_{\mathrm{S}}(\mathbf{x})$, respectively. Fig. 1 illustrates that the attribute vectors are rich enough to warrant differentiation of different anatomical structures. For robust correspondence matching, rather than computing the similarity on a voxel-to-voxel basis, we compare the similarity of the voxels in the neighborhood of $\mathbf{x}_t$ with that of $\mathbf{x}_s$, and vice-versa. For a voxel $\mathbf{x}$ in the neighborhood of $\mathbf{x}_t$ in the template space, the distance $\mathbf{x} - \mathbf{x}_t$ corresponds to $\mathit{\Delta}_s(\mathbf{x}, \mathbf{x}_t) = f(\mathbf{x}) - f(\mathbf{x}_t)$ in the subject space. Likewise, for voxels $\mathbf{x}$ and $\mathbf{x}_s$ in the subject space, their corresponding distance in the template space is $\mathit{\Delta}_t(\mathbf{x}, \mathbf{x}_s) = f^{-1}(\mathbf{x}) - f^{-1}(\mathbf{x}_s)$. We can hence define the regional similarity measures, in both forward and reverse directions, as:

$$\mathrm{sim}_{\mathrm{T}\to\mathrm{S}}(\mathbf{x}_t, \mathbf{x}_s) = m\big(\mathbf{a}_{\mathrm{T}}(\mathbf{x}_t), \mathbf{a}_{\mathrm{S}}(\mathbf{x}_s)\big) \sum_{\mathbf{x}\in\mathcal{N}_1(\mathbf{x}_t)} \frac{m\Big(\mathbf{a}_{\mathrm{T}}\big(\mathbf{x}\big), \mathbf{a}_{\mathrm{S}}\big(\mathbf{x}_s + \mathit{\Delta}_s(\mathbf{x}, \mathbf{x}_t)\big)\Big)}{|\mathcal{M}_1(\mathbf{x}_t)|},$$

$$\mathrm{sim}_{\mathrm{S}\to\mathrm{T}}(\mathbf{x}_t, \mathbf{x}_s) = m\big(\mathbf{a}_{\mathrm{T}}(\mathbf{x}_t), \mathbf{a}_{\mathrm{S}}(\mathbf{x}_s)\big) \sum_{\mathbf{x}\in\mathcal{N}_2(\mathbf{x}_s)} \frac{m\Big(\mathbf{a}_{\mathrm{T}}\big(\mathbf{x}_t + \mathit{\Delta}_t(\mathbf{x}, \mathbf{x}_s)\big), \mathbf{a}_{\mathrm{S}}\big(\mathbf{x}\big)\Big)}{|\mathcal{M}_2(\mathbf{x}_s)|},$$

where $\mathcal{N}_1(\cdot)$ and $\mathcal{N}_2(\cdot)$ denote the respective neighborhoods, and $|\mathcal{N}(\mathbf{x})|$ the cardinality of set $\mathcal{N}(\mathbf{x})$. These regional similarity measures are used to optimize the ASSAP energy function (1), which will be discussed in Section 2.3.

2.2 Landmark Selection

Brain medical images are inherently high dimensional and computations involved in registering these images can be prohibitive. In order to overcome this, out of all possible voxels $\mathbf{X}_{\mathrm{T}} = \{\mathbf{x}_t : t = 1, \ldots, M\}$ for the template, and $\mathbf{X}_{\mathrm{S}} = \{\mathbf{x}_s : s = 1, \ldots, N\}$ for the subject, $(\mathbf{X}_{\mathrm{T}}, \mathbf{X}_{\mathrm{S}} \in \mathcal{R}^3)$, we select as landmarks a subset of voxels with more distinctive attribute vector $\mathbf{X}_{\mathrm{S}}(k) = \{\mathbf{x}_t(k) :$

$t = 1, \ldots, M(k)\} \subset \mathbf{X}_\mathrm{S}$ and $\mathbf{X}_\mathrm{T}(k) = \{\mathbf{x}_s(k) : s = 1, \ldots, N(k)\} \subset \mathbf{X}_\mathrm{T}$, at iteration k of the registration, for correspondence matching. As the registration progresses, an increasing number of landmarks are used to refine the registration. Besides the obvious benefit of making the correspondence matching problem more feasible, this approach also helps mitigate the problem of local minima. Selecting an initial smaller number of landmarks essentially means that we are now solving a lower-degrees-of-freedom approximation of the transformation, and is hence less prone to be trapped by local minima. As more landmarks start to participate, transformation of increasing complexity can be estimated. F-TIMER selects as landmarks a combination of voxels with the highest tensor edge magnitudes ($H_r^{\mathrm{Tensor}}(\mathbf{x}) > \alpha_r^{\mathrm{TsrEdge}}$), FA map edge magnitudes ($H_r^{\mathrm{FA}}(\mathbf{x}) > \alpha_r^{\mathrm{FAEdge}}$) and FA values ($\mathrm{FA}_r(\mathbf{x}) > \alpha_r^{\mathrm{FA}}$), since these voxels represent important anatomical structures, and are relatively easy to locate in images with sufficient contrast. The values of $\alpha_r^{\mathrm{TsrEdge}}$, $\alpha_r^{\mathrm{FAEdge}}$ and α_r^{FA} are initially high, but are progressively lowered to allow more voxels to participate in correspondence matching as registration progresses. Fig. 2 gives an illustration of the landmarks at two different stages of the registration.

2.3 Correspondence Matching and Transformation Estimation

The determination of the non-rigid spatial mapping can be cast into an *Attributive Symmetric Soft Assignment Problem* (ASSAP). We adopt a hierarchical strategy in minimizing the related energy function. At each iteration k, the active landmarks consist of a subset of voxels $\mathbf{X}_\mathrm{T}(k) \subset \mathbf{X}_\mathrm{T}$ from the template and in $\mathbf{X}_\mathrm{S}(k) \subset \mathbf{X}_\mathrm{S}$ from the subject. Let f denote a non-rigid spatial mapping. Our goal is to find the optimal correspondence matrices $\hat{\mathbf{P}}(k)$ and $\hat{\mathbf{Q}}(k)$ and an optimal spatial transformation $\hat{f}$ by minimizing an energy function with constituent terms explained as follows:

- We would like to match the driving voxels in $\mathbf{X}_\mathrm{T}(k)$ and $\mathbf{X}_\mathrm{S}(k)$ as closely as possible in the subject space, but at the same time encourage matching of voxels with similar attribute vectors. And naturally, a voxel pair $\mathbf{x}_t(k)$ and $\mathbf{x}_s(k)$ satisfying these conditions will be deemed a more probable match and given a higher probability $p_{t,s}$ in the energy function. This can be realized with: $E(\mathbf{P}(k), f) = \sum_{t=1}^{M(k)} \sum_{s=1}^{N(k)} p_{t,s}(k) \Big\{ ||f(\mathbf{x}_t(k)) - \mathbf{x}_s(k)||^2 - \log \left[\mathrm{sim}_{\mathrm{T}\to\mathrm{S}}(\mathbf{x}_t(k), \mathbf{x}_s(k))\right] \Big\}$, where $\mathbf{P}(k) = \{p_{t,s}(k)\}$.
- We require mapping f to be consistent in that equal consideration is given to correspondence matching from the point of view of the template and the subject in determining f. This is to avoid bias towards the template or the subject, and to avoid local minima resulting from correspondence ambiguity. We thus incorporate a term in the energy function, which is symmetric to that defined above, by requiring similar matching in the template space: $E(\mathbf{Q}(k), f) = \sum_{t=1}^{M(k)} \sum_{s=1}^{N(k)} q_{t,s}(k) \Big\{ ||\mathbf{x}_t(k) - f^{-1}(\mathbf{x}_s(k))||^2 - \log \left[\mathrm{sim}_{\mathrm{S}\to\mathrm{T}}(\mathbf{x}_t(k), \mathbf{x}_s(k))\right] \Big\}$ where $\mathbf{Q}(k) = \{q_{t,s}(k)\}$.

- Soft correspondence are permitted in the initial stages of the registration so that the energy function is smooth and better behaved. Towards the end of the registration, more exact one-to-one correspondence is enforced. This is realized by energy terms: $E(\mathbf{P}(k)) = \alpha \sum_{t=1}^{M(k)} \sum_{s=1}^{N(k)} p_{t,s}(k) \log(p_{t,s}(k))$, $E(\mathbf{Q}(k)) = \alpha \sum_{t=1}^{M(k)} \sum_{s=1}^{N(k)} q_{t,s}(k) \log(q_{t,s}(k))$. Parameters α controls the degree of fuzziness of the matching. It has initially high values, encouraging fuzzy matching, and later progressively lower values, which enforce exact matching.
- Mapping f is required to be smooth so as to preserve the topology and to avoid arbitrary mapping of non-landmark voxels. This is enforced by energy term: $E(f) = \beta ||\mathcal{L}f||^2$. $\mathcal{L}$ is an operator which aids in measuring the bending energy. β is a weighting factor which is decreased throughout the registration to allow f to model deformation of increasing complexity.

The overall energy function that needs to be optimized is:

$$E(\mathbf{P}(k), \mathbf{Q}(k), f) = \big\{E(\mathbf{P}(k), f) + E(\mathbf{Q}(k), f) + E(\mathbf{P}(k)) + E(\mathbf{Q}(k)) + E(f)\big\}. \quad (1)$$

In the following, we describe an optimization strategy to minimize (1), which basically alternates between correspondence matching (**Step 1**) and dense transformation estimation (**Step 2**). We first fix f and solve for the $\mathbf{P}$ and $\mathbf{Q}$ by letting $\partial E(\mathbf{P}(k), \mathbf{Q}(k), f)/\partial p_{t,s}(k) = 0$, $\partial E(\mathbf{P}(k), \mathbf{Q}(k), f)/\partial q_{t,s}(k) = 0$. We then fix $\mathbf{P}(k)$ and $\mathbf{Q}(k)$, and solve for f using TPS.

Step 1: Correspondence Matching: The correspondence matrices $\mathbf{P}(k)$ and $\mathbf{Q}(k)$ can be updated as follows:

$$p_{t,s}(k) = \mathrm{sim}_{\mathrm{T}\to\mathrm{S}}(\mathbf{x}_t(k), \mathbf{x}_s(k)) e^{-||f(\mathbf{x}_t(k)) - \mathbf{x}_s(k)||^2/\alpha}, \; p_{t,s}(k) \leftarrow p_{t,s}(k) / \sum_{s=1}^{N(k)} p_{t,s}(k)$$

$$q_{t,s}(k) = \mathrm{sim}_{\mathrm{S}\to\mathrm{T}}(\mathbf{x}_t(k), \mathbf{x}_s(k)) e^{-||\mathbf{x}_t(k) - f^{-1}(\mathbf{x}_s(k))||^2/\alpha}, \; q_{t,s}(k) \leftarrow q_{t,s}(k) / \sum_{t=1}^{M(k)} q_{t,s}(k).$$

Overly weak matches are prevented by setting $p_{ts}(k)$ or $q_{ts}(k)$ falling below a predefined threshold to zero prior to normalization.

Step 2: Dense Transformation Estimation:

$$\sum_{t=1}^{M(k)} \sum_{s=1}^{N(k)} \Big[p_{t,s}(k) ||f(\mathbf{x}_t(k)) - \mathbf{x}_s(k)||^2 + q_{t,s}(k) ||\mathbf{x}_t(k) - f^{-1}(\mathbf{x}_s(k))||^2 \Big] + \beta ||\mathcal{L}f||^2$$

which, as it stands, is very cumbersome. A slightly different form is implemented:

$$\min_f E(f) = \min_f \left\{ \sum_{t=1}^{M(k)} ||f(\mathbf{x}_t(k)) - \hat{\mathbf{x}}_t(k)||^2 + \sum_{s=1}^{N(k)} ||f(\hat{\mathbf{x}}_s(k)) - \mathbf{x}_s(k)||^2 + \beta ||\mathcal{L}f||^2 \right\}$$

where $\hat{\mathbf{x}}_t(k) = \sum_{s=1}^{N(k)} p_{t,s}(k) \mathbf{x}_s(k)$, $\hat{\mathbf{x}}_s(k) = \sum_{t=1}^{M(k)} q_{t,s}(k) \mathbf{x}_t(k)$. The variables $\hat{\mathbf{x}}_t(k)$ and $\hat{\mathbf{x}}_s(k)$ can be seen as the newly estimated locations in the subject and

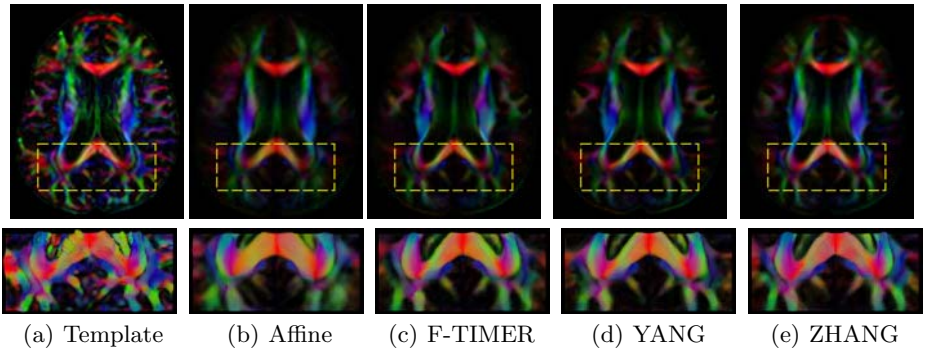
template spaces, which correspond to $\mathbf{x}_t(k)$ and $\mathbf{x}_s(k)$, respectively. The above equation can be solved using TPS fitting [5], which essentially minimizes the geometrical distance of the driving voxels and, at the same time, minimizes the bending energy $||\mathcal{L}f||^2$.

3 Experimental Results

The dataset consisted of diffusion tensor images of 22 subjects, acquired using a 1.5T scanner. The imaging dimension and resolution were 256×256 and $0.9375 \times 0.9375 \times 2.5\text{mm}^3$, respectively. Whenever appropriate, results yielded by the deformable registration algorithms, presented by Yang et al. [1] and Zhang et al. [2], will be included for comparison. They will be referred to as YANG and ZHANG, respectively, in the rest of the paper. It is worth noting here that registration using F-TIMER takes approximately 15 min on a 2.66GHz Linux machine, compared to 210 min using YANG, and 60 min using ZHANG.

3.1 Real Subjects

One subject was selected from the dataset and taken as the template. 21 subjects were then registered onto this template. By averaging all the registered images, we could visually inspect the accuracy of the registration. The result is shown in Fig. 3. It can be observed that for the FA map based affine registered images, their average image, shown in Fig. 3(b), is fuzzy especially in areas near the cortical region. In comparison, after registration with F-TIMER, the average image, Fig. 3(c), shows much improved sharpness. It is difficult to gauge the performance of F-TIMER by visual comparison with the average images yielded by YANG and ZHANG, shown in Fig. 3(c) & (d). We describe a quantitative approach in the upcoming section.



(a) Template (b) Affine (c) F-TIMER (d) YANG (e) ZHANG

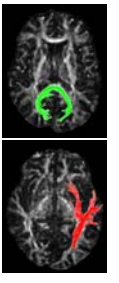
Fig. 3. Group-averaged images resulting from the registration of the 21 subjects. The FA weighted first principal directions are shown in their color coded representations: green for the anterior-posterior direction, blue for the superior-inferior direction, and red for the left-right direction. The tensors in the yellow boxes are shown in their FA weighted ellipsoidal representations in the bottom panels.

Table 1. Deformation estimation errors (mm)

	Whole Brain			Cortical Region		
	F-TIMER	YANG	ZHANG	F-TIMER	YANG	ZHANG
Mean	1.36	1.87	1.44	1.43	2.03	1.68
S. D.	1.14	1.48	1.75	1.17	1.59	1.96
p-value		3.50×10^{-8}	2.80×10^{-1}		7.71×10^{-8}	1.28×10^{-2}

Table 2. Fiber bundle distances (mm)

	F-TIMER	YANG	ZHANG
Splenium	0.87	0.91	0.98
S. D.	0.015	0.016	0.044
p-value		3.51×10^{-1}	5.06×10^{-2}
Cortical	1.34	1.49	1.91
S. D.	0.150	0.115	0.208
p-value		2.08×10^{-1}	1.30×10^{-4}



3.2 Simulated Subjects

To further evaluate the accuracy of F-TIMER, we generated 20 simulated deformation fields using the statistical model of deformation (SMD) proposed in [9]. One human brain was chosen from the dataset as the template and the 20 simulated deformation fields, which also served as the ground truths, were applied to the template, resulting in 20 simulated human brains. These 20 simulated brains were then registered back onto the template using F-TIMER and the deformation fields estimated by the registration were compared with the ground truths, using Euclidean distance. Results shown in Table 1 indicate that F-TIMER yields higher accuracy. F-TIMER is also more consistent as indicated by the smaller deviation values. Also shown in Table 1 are the results for the cortical region and similar conclusion can be drawn. The significance of the improvement of F-TIMER over YANG and ZHANG is indicated by the paired t-test p-values in the same table.

3.3 Fiber Tracking

Using a tractography method know as FACT [10], fiber bundles passing through some regions of interest (ROIs) were tracked, extracted, and compared for quantifying the registration accuracy of F-TIMER. Based on the simulated data generated above, we present here two sets of results. One ROI was selected so a fiber bundle residing in the splenium of the corpus callosum (CC) could be extracted for comparison. The second ROI was selected so that we could evaluate F-TIMER in a more difficult situation where a fiber bundle near the cortical surface was extracted. The fiber bundles are shown in Table 2. The distance of two fiber bundles was then measured using the *mean of the closest distances*, in

a way similar to that used in [2]. A summary of the results of all fiber bundles is shown in Table 2. The results indicate that F-TIMER yields better performance when compared to YANG and ZHANG.

4 Conclusion

F-TIMER extracts distinctive features from a diffusion tensor image, drawing on multiscale tensor regional and boundary information, as automated structural landmarks. Employing these landmarks to minimize an *Attributive Symmetric Soft Assignment Problem* (ASSAP) related energy function, a smooth, topology preserving, and consistent transformation can be derived. Emperimental results show that F-TIMER can achieve sufficiently good accuracy with relatively low computation cost.

Acknowledgments. This work was supported in part by grant 1R01EB006733.

References

1. Yang, J., Shen, D., Davatzikos, C., Verma, R.: Diffusion tensor image registration using tensor geometry and orientation features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 905–913. Springer, Heidelberg (2008)
2. Zhang, H., Yushkevich, P.A., Alexander, D.C., Gee, J.C.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. Med. Image Anal. 10(5), 764–785 (2006)
3. Shen, D., Davatzikos, C.: HAMMER: Heirarchical attribute matching mechanism for elastic registration. IEEE TMI 21(11), 1421–1439 (2002)
4. Christensen, G.E., Johnson, H.J.: Consistent image registration. IEEE TMI 20(7), 568–582 (2001)
5. Bookstein, F.L.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE TPAMI 11(6), 567–585 (1989)
6. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. CVIU 89(2-3), 114–141 (2003)
7. Xu, D., Mori, S., Shen, D., van Zijl, P.C.M., Davatzikos, C.: Spatial normalization of diffusion tensor fields. Mag. Res. in Med. 50(1), 175–182 (2003)
8. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-Euclidean metrics for fast and simple calculus on diffusion tensors. Mag. Res. in Med. 56(2), 411–421 (2006)
9. Xue, Z., Shen, D., Karacali, B., Stern, J., Rottenberg, D., Davatzikos, C.: Simulating deformations of MR brain images for validation of atlas-based segmentation and registration algorithms. NeuroImage 33(3), 855–866 (2006)
10. Jiang, H., van Zijl, P.C., Kim, J., Pearlson, G.D., Mori, S.: DtiStudio: resource program for diffusion tensor computation and fiber bundle tracking. Computer Methods and Programs in Biomedicine 81(2), 106–116 (2006)

DISCO: A Coherent Diffeomorphic Framework for Brain Registration under Exhaustive Sulcal Constraints

Guillaume Auzias[1], Joan Glaunès[2], Olivier Colliot[1], Matthieu Perrot[3], Jean-François Mangin[3], Alain Trouvé[4], and Sylvain Baillet[1,5]

[1] Université Pierre et Marie Curie-Paris 6, UMR 7225, UMRS 975, Centre de Recherche de l'Institut Cerveau-Moelle (CRICM), Paris, France
[2] MAP5, Université Paris 5 - René Descartes, Paris, France
[3] NeuroSpin, CEA, Orsay, France
[4] CMLA, ENS de Cachan, France
[5] Medical College of Wisconsin, Milwaukee, USA
guillaume.auzias@upmc.fr

Abstract. Neuroimaging at the group level requires spatial normalization of individual structural data. We propose a geometric approach that consists in matching a series of cortical surfaces through diffeomorphic registration of their sulcal imprints. The resulting 3D transforms naturally extends to the entire MRI volumes. The DIffeomorphic Sulcal-based COrtical (DISCO) registration integrates two recent technical outcomes: 1) the automatic extraction, identification and simplification of numerous sulci from T1-weighted MRI data series hereby revealing the sulcal imprint and 2) the measure-based diffeomorphic registration of those crucial anatomical landmarks. We show how the DISCO registration may be used to elaborate a sulcal template which optimizes the distribution of constraints over the entire cortical ribbon. DISCO was evaluated through a group of 20 individual brains. Quantitative and qualitative indices attest how this approach may improve both alignment of sulcal folds and overlay of gray and white matter volumes at the group level.

1 Introduction

Correct alignment of cortical surfaces amongst a group of subjects is crucial in neuroimaging because of the large variety of breakthrough applications. Intensity-based approaches cannot achieve this issue as the matching criteria being used is global and not optimized for finer-scale warping at the cortical level. In that respect, research in volume-based methods has evolved from the mere alignement of intesity-based elements to the integration of geometrical features of the cortical surface and even recently of cortical surface itself with hybrid surfacic and volumic approaches [9,12]. However, as in more standard methods like [2], the resulting deformation constraints enforce the alignment of cortical valleys and crests but without any guarantee that sulci of identical anatomical denomination would be properly aligned altogether (e.g. the precentral sulcus from a

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 730–738, 2009.

subset of subjects might end up being matched to the central sulcus of other subjects). Specific alignment of cortical circonvolutions remains crucial as the sulcal folds shape might be correlated with the underlying functional organization [4] and requires higher level anatomical knowledge. Indeed, explicit sulcal constraints have been experimented in e.g. [14,7,13]. Joshi et al. have proposed to combine a surfacic sulcal warp with a subsequent volumic matching step [8]. However, in all those approaches the similarity between sulcal landmarks relies on their parametrisation which impose strong assumptions about their topology and therefore about the segmentation. Few sulcal landmarks are then selected manually; which results in a sparse and heterogeneous distribution of anatomical landmarks guiding the transformations. Heterogeneities in the constraints favors irregularities of deformation fields and might result in alterations of the overall topology of the cortex.

In this contribution, we optimize the alignment of cerebral structures among a group of subjects through the registration of their exhaustive and sulcus-based folding patterns while constraining the dense 3D transform to be diffeomorphic – that is, smooth and invertible – through Large Deformation Diffeomorphic Metric Mapping (LDDMM) framework [5]. The DISCO registration technique – DIffeomophic Sulcal-based COrtical registration – consists in the following steps:

1. The automatic extraction, identification and simplification of up to 120 sulci per subject. This step yields a dense set of distributed landmarks, denoted as the individual *sulcal imprint*, and prevents the use of tedious and to a great extent subjective, manual tracing.
2. An LDDMM transform aiming at matching individual sulcal imprints obtained from Step 1. The similarity between corresponding sulci rely on measure based-representation which avoids the parametrisation of landmarks. We describe how this procedure can be derived from a group of subjects of arbitrary size through the definition of an adaptive sulcal template derived from the very set of subjects involved. Note that the LDDMM framework naturally yields the ability to extend the deformation to any additional object in 3D – such as deeper brain structures – thereby overcoming the major limitation of surface-based approaches.

The performances of DISCO are quantitatively evaluated using data from a group of 20 healthy control subjects.

2 Extraction of the Individual Sulcal Imprint

The extraction of the individual sulcal imprint from T1-weighted MR image volumes is initiated by the automatic segmentation and labeling of a large number of sulci using the brainVISA free software platform [1,10]. Voxel labeling of the gray and white matter tissues are obtained from histogram analyses and mathematical morphological techniques applied to the biased-corrected T1-weighted MR images. Elementary sulcal elements are segmented and divided into topologically-simple surfaces and organized as a graph structure. The sulci are

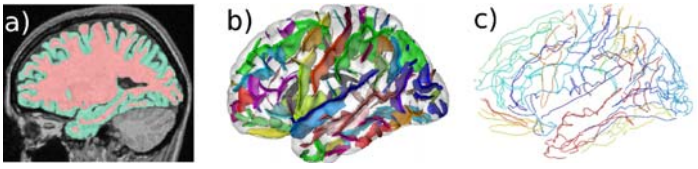


Fig. 1. Automatic extraction of the sulcal imprint: a) tissue classification; b) extraction of sulci and automatic identification results in up to 60 individual sulcal labels in each hemisphere. c) The automatic simplification of the original complex sulcal pattern yields a distributed set of sulcal edges, thereby defining the *sulcal imprint* of an individual brain.

then automatically labeled according to a predefined nomenclature of 60-sulcus labels per hemisphere as illustrated on Fig.1 b). Agreement between the computer and human experts reaches 86% on average [11]. Though the performances of the automatic labeling procedure might reach up to 96% for well-defined folds such as the central sulcus, they tend to decline in regions where the cortical folding pattern has shows considerable inter-individual variability such as in the occipital areas.

This extraction yields structures with complex shapes made of subsets of voxels corresponding to an overly detailed description of the sulci in the context of inter-individual registration, while it is crucial to preserve the variability in topology of a given sulcus across individuals. The original complex sulcal structures are subsequently and automatically simplified as relevant sulcal landmarks using the original procedure detailed thereafter.

2.1 From Sulci to Sulcal Imprint

For each sulcus s, the initial set of voxels is first reduced to two subsets: one corresponding to the *fundus* of s, defined as the border of the sulcus reaching the deepest into the brain volumes, and one for its *outer border*, defined as the junction between the sulcus and the hull of the brain. These voxels are readily identified during the very sulcal extraction process as performed by brainVISA.

These two initial subsets are then independently reduced and smoothed using K-means clustering as illustrate Fig.2b). The voxels are geometrically grouped into smaller clusters which are subsequently reduced to their respective centroid. The resulting points along either sulcal border are then connected using the minimum spanning tree approach, thereby yielding two open connected graph structures, i.e. two trees (see Fig.2c). Finally, the secondary branches of the trees are removed using a longest-path approach (Fig.2e). This way, the fundus and outer borders of each sulcus s are reduced to two edges denoted as E_s^f and E_s^o respectively. The extraction of the sulcal edges associated to an interrupted sulcus remains identical. The sulcus will be described as two connected graphs through the minimum spanning tree. However, the interruptions will appear as holes in the distribution of points along the respective path. This procedure yields a set of folding features distributed across the entire cortical surface as

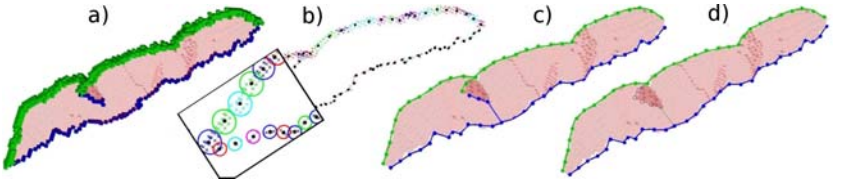


Fig. 2. a) Each sulcus is decomposed into three subsets of voxels: fundus (in blue), outer edge (green) and other voxels (in red) using brainVISA. The description of each sulcus may be summarized by fundus and outer border voxel subsets. b) Those voxels are grouped into clusters (shown here as colored circles) and each cluster is reduced to its barycenter (black dots). Finally, the sulcal borders illustrated in c) are reduced to simple lines in d) through a longest path approach.

illustrated Fig. 1 c). These features hereby define the *sulcal imprint* of each individual brain anatomy – $\mathcal{I} = [E_1^f, E_1^o, .., E_s^f, E_s^o, .., E_S^f, E_S^o]$ – that will be matched across subjects.

3 Measure-Based Diffeomorphic Matching of Sulcal Imprints

We introduce a non-linear pairwise registration approach of sulcal imprints in the general framework of the LDDMM theory [5]. The deformation ϕ of sulcal edge E_1 onto another sulcal edge E_2 is defined as the minimum of the following registration energy functional:

$$J^{sulc}_{E_1,E_2}(\phi) = \gamma Reg(\phi) + Mis(\phi(E_1), E_2), \tag{1}$$

where the first term controls for the regularity of the deformation while the second term evaluates the mismatch between the deformed sulcal edge $\phi(E_1)$ and the target E_2; γ being a scalar trade-off parameter. Following the LDDMM theory, ϕ is a diffeomorphism if it may be defined as a solution at time $t = 1$ of the differential equation: $\partial_t \phi^{\mathbf{v}}_t = \mathbf{v}_t \circ \phi^{\mathbf{v}}_t$, with initial condition $\phi^{\mathbf{v}}_0 = Id$. Id represents the identity deformation that maps an object onto itself. In this equation, $\mathbf{v}_t : \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ is a time-dependent vector field which models the infinitesimal variations of the deformation flow. $\mathbf{v}_t$ belongs to the reproducing kernel Hilbert space V of regular vector fields. V is associated to a kernel K^V controlling for the regularity of the final diffeomorphic transform. We define the cost of a given diffeomorphism ϕ as its distance to the identity transform: $d^2_V(Id, \phi) = \inf_{\mathbf{v}} \left\{ \int_0^1 ||\mathbf{v}_t||^2_V dt, \phi^{\mathbf{v}}_1 = \phi \right\}$. Registering a pair of sulcal edges in a diffeomorphic framework then consists in minimizing the functional:

$$J^{sulc}_{E_1,E_2}(\phi) = \gamma d^2_V(id, \phi) + Mis(\phi(E_1), E_2). \tag{2}$$

The mismatch Mis between two sulcal edges is defined hereafter.

3.1 Description of Sulcal Landmarks as Measures

Each couple of anatomically-corresponding sulcal edges are considered as two sets of points $E_1 = (x_i)_{i<n_x}$ *and* $E_2 = (y_j)_{j<n_y} \subset \mathbb{R}^3$, with possibly $n_x \neq n_y$. These two sets of points can be described mathematically as measures μ and ν respectively, each consisting of the weighted sum of Dirac distributions [5]: $\mu = \sum_{i=1}^{n_x} a_i \delta_{x_i}$ and $\nu = \sum_{i=1}^{n_y} b_i \delta_{y_i}$. $(a_i)_{i<n_x}$ and $(b_j)_{j<n_y}$ are two sets of scalar weight parameters. These weights are set as follows: if the neighborhood of each x_i in the associated sulcal edge is denoted as H, then $a_i = \frac{1}{card(H)} \sum_{h|h\in H} ||x_h - x_i||_{\mathbb{R}^3}$. This uniformly distributes the weights along the measure, thereby compensating for the heterogeneity in the spatial distribution of points along the edge lines corresponding to interrupted sulci. The action of ϕ on the measure μ may be defined as a mass transportation problem: $\phi(\mu) = \phi(\sum_i a_i \delta_{x_i}) = \sum_i a_i \delta_{\phi(x_i)}$. In order to evaluate the adjustment of the source sulcal measure μ to the target measure ν, we introduce an additional reproducing kernel Hilbert space I associated to a second kernel K^I, such that every bounded and signed measure belongs to I^*, the dual space of I. We may then evaluate the distance between the pair of measures μ and ν as: $d_I^2(\phi(\mu), \nu) = ||\phi(\mu) - \nu||_{I^*}^2 = \sum_{i,j} a_i a_j K^I(\phi(x_i), \phi(x_j)) + \sum_{i,j} b_i b_j K^I(y_i, y_j) - 2\sum_{i,j} a_i b_j K^I(\phi(x_i), y_j)$. Equation (2) can therefore be rewritten as:

$$J_{\mu,\nu}^{sulc}(\phi) = \gamma d_V^2(id, \phi) + ||\phi(\mu) - \nu||_{I^*}^2. \tag{3}$$

Considering now two sulcal imprints with P sulcal edges in common, we define $\mathcal{I}_1 = [\mu^1, .., \mu^P]$, the source sulcal imprint to be adjusted to a target sulcal imprint $\mathcal{I}_2 = [\nu^1, .., \nu^P]$. Registering a pair of brains through their respective sulcal imprints corresponds to the minimization of the following functional:

$$J_{\mathcal{I}_1,\mathcal{I}_2}^{impr}(\phi) = \gamma d_V^2(id, \phi) + \sum_{p=1}^{P} ||\phi(\mu^p) - \nu^p||_{I^*}^2. \tag{4}$$

Of primary importance is that the resulting deformation is a fully 3D diffeomorphic map defined everywhere in $\mathbb{R}^3$, hence not only on the cortical surface, but also in the entire MRI volume.

3.2 Unbiased, Empirical Anatomical Template for Multiple-Subject Registration

We now propose a multi-scale iterative approach at the group level that extends previous results from [6]. This approach avoids the arbitrary selection of a single brain as a registration template while exploiting the maximum of the sulcal information available. The methodology involved is itemized as follows:

1. The sulcal imprint is extracted from each individual brain data and is linearly registered into standardized Talairach space [3].
2. An empirical template is defined as the union of the entire set of sulcal points through the entire group of subjects involved in the study. For each sulcal

label available across the group, the corresponding sulcal landmark in the template consists of the union of all points associated to this label within every subject of the group.
3. Diffeomorphic transformation of each individual data onto the empirical template is operated following the methodology described in Section 3.
4. The process in steps 2 and 3 is iterated Q times by considering the resulting transformed sulcal points as the new running template, until the evolution between two successive template samples is inferior to a fixed threshold. We focus on finer scale deformations as the template is iteratively refined by reducing the sizes of the kernels involved in the registration energy (for both regularization and mismatch terms in Eq. (5)) between two iterations.

Therefore, registering N subjects consists of Q iterations of N minimization steps while the empirical template is updated between two iterations, in a multi-scale framework.

For clarity purposes, we now detail the elaboration of an empirical template consisting of a single sulcus through a group of N individual sulcal imprints. The generalization to P sulcal labels extends the strategy introduced by Eq. (2) through Eq. (4). Following [6], let us denote $(x_{ip})_{1\leq i\leq N, 1\leq p\leq n_i}$, the N individual sets consisting of n_i points describing the sulcus to be matched across subjects; $a_{ip} \in \mathbb{R}$ their associated weights and $\mu_i = \sum_{p=1}^{n_i} a_{ip}\delta_{x_ip}$, their respective measure form. Obtaining the measure μ of the group template may be defined as a minimisation problem: $\{\hat{\phi_i}, \hat{\mu}\} = \arg\min_{\phi_i,\mu} \sum_{i=1}^{N} \left\{\gamma d_V^2(id, \phi_i) + ||\phi_i(\mu_i) - \mu||_{I^*}^2\right\}$. Note that for fixed ϕ_i, $\hat{\mu}$ reduces to the sum of the Dirac masses associated to the union of all points $\phi_i(x_{ip})$:$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \phi_i(\mu_i) = \frac{1}{N}\sum_{i=1}^{N}\sum_{p=1}^{n_i} a_{ip}\delta_{\phi_i(x_{ip})}$. Hence, the problem reduces to:

$$\{\hat{\phi_i}\} = \arg\min_{\phi_i} \sum_{i=1}^{N} \left\{\gamma d_V^2(id, \phi_i) + ||\phi_i(\mu_i) - \frac{1}{N}\sum_{i=1}^{N}\phi_i(\mu_i)||_{I^*}^2\right\}. \quad (5)$$

At the end of the process, the transformation that brings subject i into the common space is the composition of Q diffeomorphisms: $\phi_i^Q \circ \phi_i^{Q-1} \circ \ldots \circ \phi_i^1$. By definition, the composition of Q such diffeomorphisms is also diffeomorphic.

4 Results

DISCO was applied for evaluation purposes to the registration of the brains from 20 healthy subjects through the elaboration of their corresponding empirical anatomical template. Sulcus automatic labeling was checked for errors manually. Between 88 and 98 sulci were identified across subjects in this group, 71 of which were common to all subjects and only 8 sulci were found in less than 14 of the subjects. The iterative sulcal template refinement process converged though $Q = 9$ iterations with template update. Computation time on a cluster of 10 processors running at 3.6GHz with 2Gb of RAM was 10 hours.

As shown on Fig.3, the superimposition of most sulcal edges is clearly improved after diffeomorphic transformation using DISCO. Further, it is important

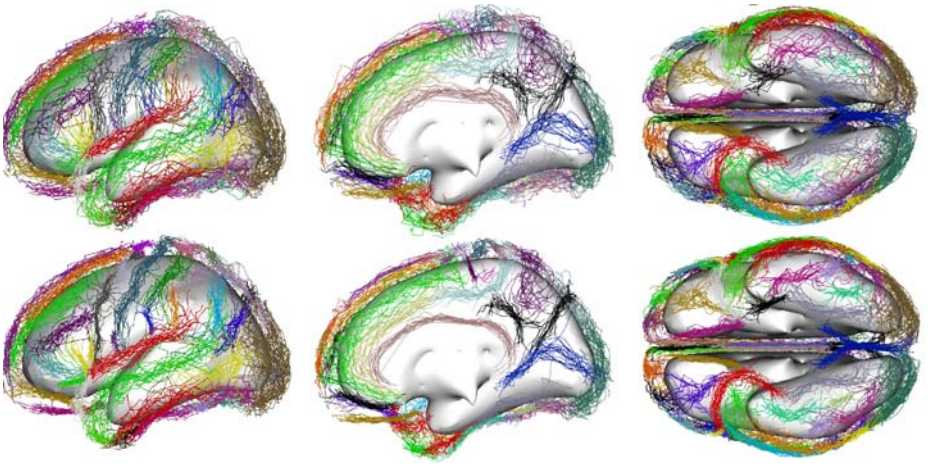


Fig. 3. Comparative results of the registration of 20 brains using the linear technique [3] (upper panel) vs. DISCO diffeomorphic sulcal matching (lower panel)

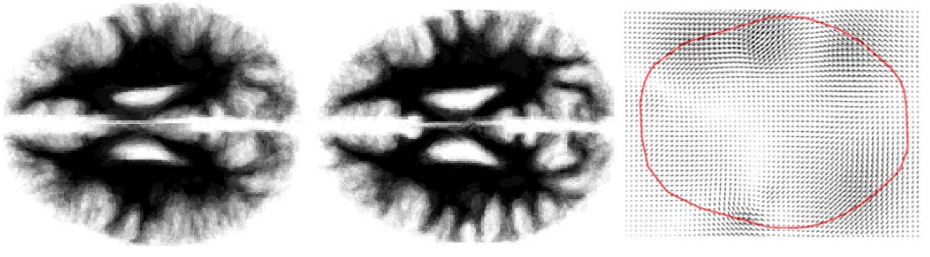


Fig. 4. Comparison of the linear (left) and DISCO (middle) registration procedures across 20 white-matter masks. Correct alignment of cortical circumvolutions would yield reduced fuzziness in the gray levels of the average volume masks, as readily observed after DISCO was applied. Right: a typical slice of the resulting DISCO deformation field. The convex hull of the brain in that particular slice is shown in red.

to remember that *matching is not warping*, as some aspects of the original group variability have been preserved by the regularity of the diffeomorphic transforms.

The Hausdorff distance between each subject and the rest of the group was computed for each sulcal landmark, and averaged across subjects, after linear or diffeomorphic matching. This measure is an indicator of the spatial dispersion of sulcal edges. The Hausdorff distance was reduced by 18% (from 17.2mm to 14mm) on average across all sulci and subjects using the diffeomorphic matching compared to the linear registration approach. For instance, the sulcal dispersion has been decreased by more than 7mm in the inferior part of the temporal lobe. The improvement was smaller in regions with sulcal objects of greater geometrical irregularity across subjects as in e.g. the occipital lobe. The dispersion of the central sulcus was reduced by "only" 1.6 mm as it had been already correctly realigned by the

linear registration procedure. Note also that the deformation of any given sulcus is tempered by the transformations occurring in its neighborhood.

As discussed in Section 3, the DISCO transformation naturally extends to the entire 3D volume as illustrated Fig.4 through the group averages of the binary white-matter masks.

5 Conclusion

The suggested approach combines the attractive properties of diffeomorphic matching with the pairing of anatomical landmarks considered by neuroanatomical experts. Future work will focus on larger-scale validation, including comparison with other non-linear registration methods. The preliminary results presented here suggest that this technique may lead to a new systematic approach for anatomical registration in neuroimaging group studies.

References

1. BrainVISA/Anatomist website, http://brainvisa.info
2. Ashburner, J.: A fast diffeomorphic image registration algorithm. Neuroimage 38, 95–113 (2007)
3. Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C.: Automatic 3D Intersubject Registration of MR Volumetric Data in Standardized Talairach Space. J. Comp. Assist. Tomog. 18, 192–205 (1994)
4. Fischl, B., Rajendran, N., Busa, E., Augustinack, J., Hinds, O., Yeo, B.T., Mohlberg, H., Amunts, K., Zilles, K.: Cortical Folding Patterns and Predicting Cytoarchitecture. Cerebral Cortex (2007)
5. Glaunès, J., Trouvé, A., Younes, L.: Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching. In: Proc. IEEE Conf. Comp. Vis. Pat. Rec., vol. 2, pp. 712–718 (2004)
6. Glaunès, J., Joshi, S.: Template estimation form unlabeled point set data and surfaces for computational anatomy. In: MICCAI 2006. LNCS, pp. 58–65. Springer, Heidelberg (2006)
7. Hellier, P., Barillot, C.: Cooperation between local and global approaches to register brain images. In: Insana, M.F., Leahy, R.M. (eds.) IPMI 2001. LNCS, vol. 2082, pp. 315–328. Springer, Heidelberg (2001)
8. Joshi, A.A., Shattuck, D.W., Thompson, P.M., Leahy, R.M.: Surface-Constrained Volumetric Brain Registration Using Harmonic Mappings. IEEE Trans. Med. Imag. 26, 1657–1669 (2007)
9. Liu, T., Shen, D., Davatzikos, C.: Deformable registration of cortical structures via hybrid volumetric and surface warping. Neuroimage 22, 1790–1801 (2004)
10. Mangin, J.-F., Rivière, D., Cachia, A., Duchesnay, E., Cointepas, Y., Papadopoulos-Orfanos, D., Scifo, P., Ochiai, T., Brunelle, F., Régis, J.: A framework to study the cortical folding patterns. Neuroimage 23, 129–138 (2004)
11. Perrot, M., Rivière, D., Tucholka, A., Mangin, J.-F.: Joint Bayesian Cortical Sulci Recognition and Spatial Normalization. In: IPMI. LNCS. Springer, Heidelberg (to appear)

12. Postelnicu, G.M., Zöllei, L., Fischl, B.: Combined Volumetric and Surface Registration. IEEE Trans. Med. Imag. 28, 508–522 (2008)
13. Qiu, A., Miller, M.I.: Cortical hemisphere registration via large deformation diffeomorphic metric curve mapping. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 186–193. Springer, Heidelberg (2007)
14. Shi, Y., Thompson, P.M., Dinov, I., Osher, S., Toga, A.W.: Direct cortical mapping via solving partial differential equations on implicit surfaces. Med. Imag. Anal. 11, 207–223 (2007)

Evaluation of Lobar Biomechanics during Respiration Using Image Registration

Kai Ding[1], Youbing Yin[2], Kunlin Cao[3], Gary E. Christensen[3], Ching-Long Lin[2], Eric A. Hoffman[1,4], and Joseph M. Reinhardt[1,⋆]

[1] Department of Biomedical Engineering
[2] Department of Mechanical and Industrial Engineering
[3] Department of Electrical and Computer Engineering
[4] Department of Radiology
The University of Iowa, Iowa City, IA 52242
{kai-ding,youbing-yin,kunlin-cao,gary-christensen,
ching-long-lin,eric-hoffman,joe-reinhardt}@uiowa.edu

Abstract. The human lungs are divided into five independent compartments called lobes. The lobar fissures separate the lung lobes. It is hypothesized that the lobar surfaces slide against each other during respiration. We propose a method to evaluate the sliding motion of the lobar surfaces during respiration using lobe-by-lobe mass-preserving non-rigid image registration. We measure lobar sliding by evaluating the relative displacement on both sides of the fissure. The results show a superior-inferior gradient in the magnitude of lobar sliding. We compare whole-lung-based registration accuracy to lobe-by-lobe registration accuracy using vessel bifurcation landmarks.

1 Introduction

The human lungs are divided into five independent compartments called lobes. A lobar fissure is a thin space (approximately 0.5mm depending on volume of pleural fluid) separating the lung lobes. The left lung is divided into the left upper (LUL) and left lower (LLL) lobes, separated by the oblique fissure. The right lung is partitioned into the right upper lobe, middle lobe, and the lower lobe, separated by the oblique and horizontal fissures. The branching patterns of the bronchial and vascular trees also follow the lobar structure of the lung.

Regional function and biomechanics depend on the material properties of the lung parenchyma and the complex interaction between the lobes, diaphragm, and chest wall. Hubmayr et al. [1] have used embedded metal markers and X-ray projection images to study regional lung mechanics. Recently, image registration has been used to assess regional lung function and tissue biomechanics using multiple 3D images at different lung volumes by CT [2,3,4] and MRI [5]. Although those results show regional changes in lung function and mechanics, they do not explicitly account for the interaction between the lung lobes. It is believed that

⋆ J. M. Reinhardt and E. A. Hoffman are shareholders in VIDA Diagnostics, Inc.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 739–746, 2009.

during respiration the lobes move relative to each other (sliding and rotation), and this motion may provide a means to reduce the lung parenchymal distortion and avoid regions of high local stress [1]. In addition, understanding of how lobes slip relative to one another is of importance to the understanding of how the lung accommodates chest wall shape changes while minimizing effects on regional distribution of ventilation.

In this paper, we investigate lung biomechanics using a lobe-by-lobe registration technique. Our approach explicitly accounts for the registration displacement field discontinuity at the fissure (due to lobar sliding), and should provide more accurate image registration near the fissure and, as a result, better biomechanical measurements. We measure lobar sliding by evaluating the relative displacement on both sides of the fissure. We compare whole-lung-based registration accuracy to lobe-by-lobe registration accuracy using vessel landmarks.

2 Materials and Methods

2.1 Data Acquisition

All data were gathered under a protocol approved by our institutional review board. Three pairs of volumetric CT data sets from three normal human subjects were used in this study. Each image pair was acquired with a Siemens Sensation 64 multi-detector row CT scanner (Forchheim, Germany) during breath-holds near functional residual capacity (FRC) and total lung capacity (TLC) in the same scanning session. Each volumetric data set was acquired at a section spacing of $0.5 \sim 0.6$ mm and a reconstruction matrix of 512×512. In-plane pixel spacing is approximately 0.6 mm $\times$ 0.6 mm.

2.2 Automatic Lobe Segmentation

To perform the lobe-by-lobe registration, the lobes are first automatically segmented using the method from [6]. The lobar segmentation begins with automatic lung, airway tree, and vessel tree segmentation. A watershed transform, applied to a distance map derived from the original CT image and the vessel segmentation, provides an initial lobar segmentation. The lobar surfaces are refined using a 3D optimal surface detection that divides the lungs at the fissure surfaces. For complete details on the lobar segmentation method, see [6].

2.3 Image Registration

The CT scans at FRC and TLC are registered for each subject. The FRC–TLC image pairs show large lung volume change, large tissue deformation, and large voxel intensity changes. To account for these differences between the images during registration, we used a lung mass preserving registration method [7]. The method uses a similarity metric that estimates the local tissue and air fraction within the lung and minimizes the local tissue mass difference. This method has

been shown to be effective at registering across large lung volume changes (such as FRC–TLC pairs) [7].

From the CT value of a given voxel, the tissue volume can be estimated as

$$V(\mathbf{x}) = \nu(\mathbf{x}) \frac{I(\mathbf{x}) - HU_{air}}{HU_{tissue} - HU_{air}} = \nu(\mathbf{x})\gamma(I(\mathbf{x})), \tag{1}$$

where $\nu(\mathbf{x})$ denotes the volume and $I(\mathbf{x})$ is the intensity of a voxel at position $\mathbf{x}$. HU_{air} and HU_{tissue} refer to the intensity of air and tissue, respectively [8]. In this work, we assume that air is -1000 HU and tissue is 55 HU. $\gamma(I(\mathbf{x}))$ is introduced for notational simplicity.

Given (1), we can then define the similarity measure as the sum of squared local tissue volume difference:

$$C = \sum_{\mathbf{x}\in\Omega} \left[V_r(\mathbf{x}) - V_f\left(T(\mathbf{x})\right)\right]^2 = \sum_{\mathbf{x}\in\Omega} \left[\nu_r(\mathbf{x})\gamma(I_r(\mathbf{x})) - \nu_f\left(T(\mathbf{x})\right)\gamma(I_f(\mathbf{x}))\right]^2, \tag{2}$$

where Ω denotes the overlapping lung regions in the two images, and $T(\mathbf{x})$ is the warping function. In this work, $T(\mathbf{x})$ is a cubic B-splines transform:

$$T(\mathbf{x}) = \mathbf{x} + \sum_{\mathbf{k}\subset K} \phi_{\mathbf{k}}\, \beta(\mathbf{x}), \tag{3}$$

where ϕ describes the displacements of the control nodes and $\beta(\mathbf{x})$ is a three-dimensional tensor product of basis functions of cubic B-Spline.

Given a warping function $T(\mathbf{x})$, $I_f(T(\mathbf{x}))$ can be interpolated from the moving image. $\nu_f(T(\mathbf{x}))$ can be calculated from the Jacobian $J(\mathbf{x})$ of the deformation

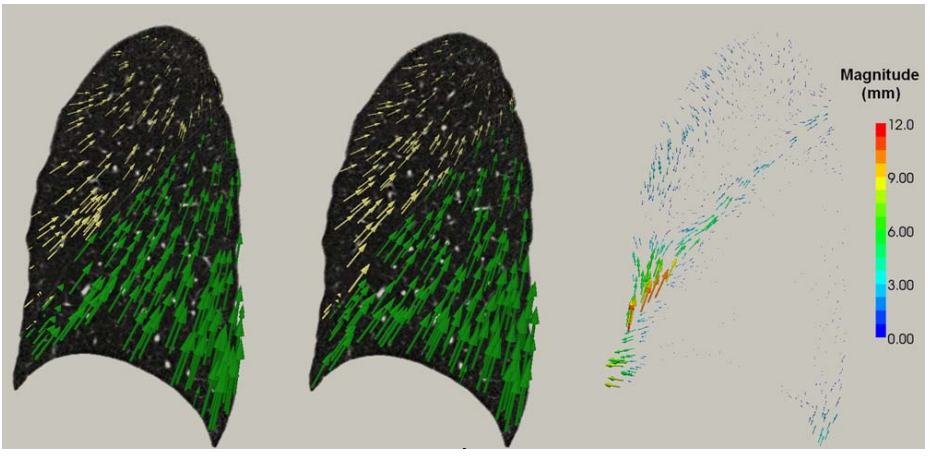


Fig. 1. Comparison of displacement field between the lobe-by-lobe registration (left column) and the lung-by-lung registration (middle column) for the LUL (yellow) and LLL (green). The right column is the difference of the two displacement fields with the magnitude indicated by the color bar.

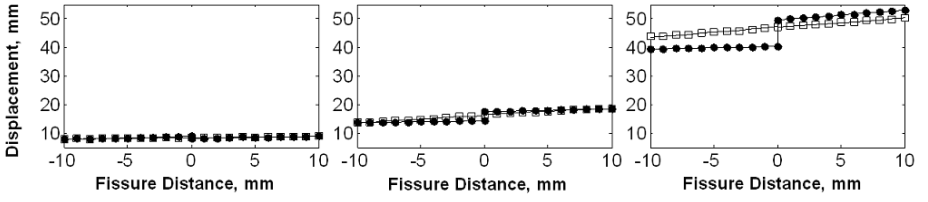
as $\nu_f(T(\mathbf{x}) = J(\mathbf{x})\nu_r(\mathbf{x})$. Note that the Jacobian value must be positive here, which can be achieved by using displacement constraints on the control nodes.

2.4 Computational Setup

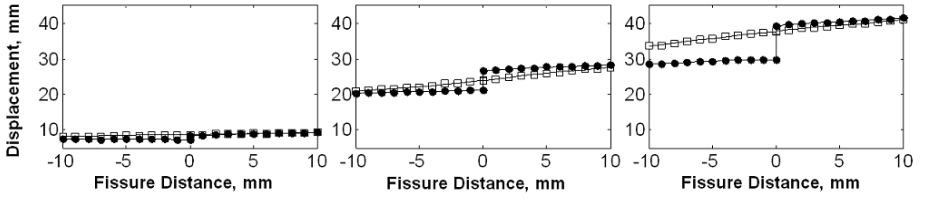
In this study, the lobe-by-lobe registration is used to investigate lobar sliding. Our current analysis is limited to the upper and lower lobes of the left lung, since the three lobes in the right lung will likely have more complicated interaction.

We start with the lobar segmentations of the TLC and FRC images as described in 2.2. After segmentation we match the TLC left upper lobe to the FRC left upper lobe, and the TLC left lower lobe to the FRC left lower lobe. After registration, the displacement fields are recombined into one left lung displacement field. We also perform conventional lung-by-lung registration to match the TLC left lung to the FRC left lung, using the same registration algorithm.

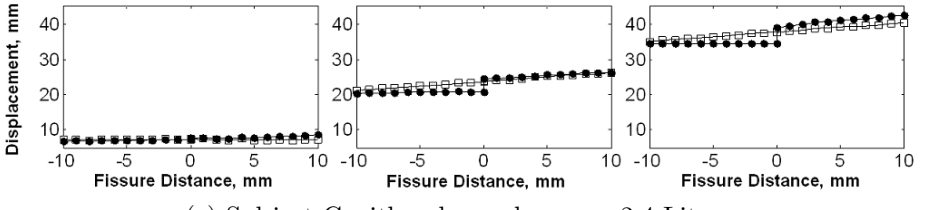
Discontinuities of the displacement field along the fissure surface are indications of lobar sliding. Figure 1 shows the displacement fields generated by lobe-by-lobe and lung-by-lung registration methods for one subject. The figure shows a considerable difference between these methods along the fissure surface.



(a) Subject A with volume change = 3.2 Liters



(b) Subject B with volume change = 2.3 Liters



(c) Subject C with volume change = 2.4 Liters

Fig. 2. Displacement profile of tangent components along a line perpendicular to the fissure surface at three different locations (left: near apex; middle: near lingula; and right: near base) for both the whole-lung-based (square) and the lobe-based (solid circle) methods

2.5 Assessment of Image Registration Accuracy

Vascular bifurcation points are used as landmarks to evaluate registration accuracy. An observer uses a landmark annotating system [9] to find corresponding landmarks in the FRC and TLC images. For each landmark, the actual landmark position is compared to the registration-derived estimate of landmark position from the two registration methods and the landmark error is calculated.

2.6 Evaluation of Local Lobar Sliding

Once the lobe segmentations are obtained, the oblique fissure surface between LUL and LLL is extracted as a triangular mesh. The normal direction is then calculated at each vertex of the mesh. The sliding motion is quantified for each point along the fissure surface by looking at the discontinuity in the line profile perpendicular to the fissure surface as shown in Figure 2. On each side, the profile of tangent component of the displacement is fitted as a 3rd order polynomial function $\mathbf{d}$ of the distance to the fissure surface. The sliding distance $s(\mathbf{x})$ at fissure surface position $\mathbf{x}$ is then defined as

$$s(\mathbf{x}) = \| \mathbf{d}_0^+ - \mathbf{d}_0^- \|, \quad \mathbf{x} \in \mathcal{S}. \tag{4}$$

where $\mathbf{d}_0^+$ is the predicted value on the fissure surface from the polynomial function along the positive normal direction (we define the normal direction pointing the LUL as positive.) while $\mathbf{d}_0^-$ is the predicted value from the other side.

3 Results

3.1 Registration Accuracy

For each lobe, 20 to 40 landmarks are identified. Table 1 shows the results of the landmark distance before and after registration for the lobe-based and whole-lung based registrations. The average landmark errors are 0.83 mm and 0.73 mm for whole-lung-based registration and lobe-based registration.

Table 1. Comparison of registration accuracy between lobe-based and whole-lung-based registrations. Distances in mm.

Subject	Lobe	Before Registration	Whole-lung-based Registration	Lobe-based registration
A	LUL	19.08 ± 8.25	0.99 ± 0.99	0.95 ± 0.81
	LLL	35.79 ± 12.69	0.94 ± 1.12	0.71 ± 0.41
B	LUL	15.09 ± 4.03	0.72 ± 0.81	0.57 ± 0.30
	LLL	38.33 ± 6.55	0.87 ± 0.48	0.75 ± 0.43
C	LUL	13.45 ± 6.59	0.78 ± 0.72	0.78 ± 0.83
	LLL	35.45 ± 10.76	0.68 ± 0.24	0.67 ± 0.30

3.2 Local Lobar Sliding

The sliding distance at each fissure surface point was calculated to evaluate the local lobar sliding. A step length of 1 mm and 10 sample points were used along the normal direction on each side of the surface. A surface point was not taken into consideration if any of its 20 sample points were outside the lobes. Figure 2 shows the displacement profile of the tangent component along a line perpendicular to the fissure surface at three different locations (near the apex, near the lingula and near the base) for the whole-lung-based (square) and the lobe-based (solid circle) methods. The results show increased sliding (larger discontinuity) in the more basal positions using the lobe-by-lobe analysis. However, these discontinuities are not apparent using the lung-by-lung analysis.

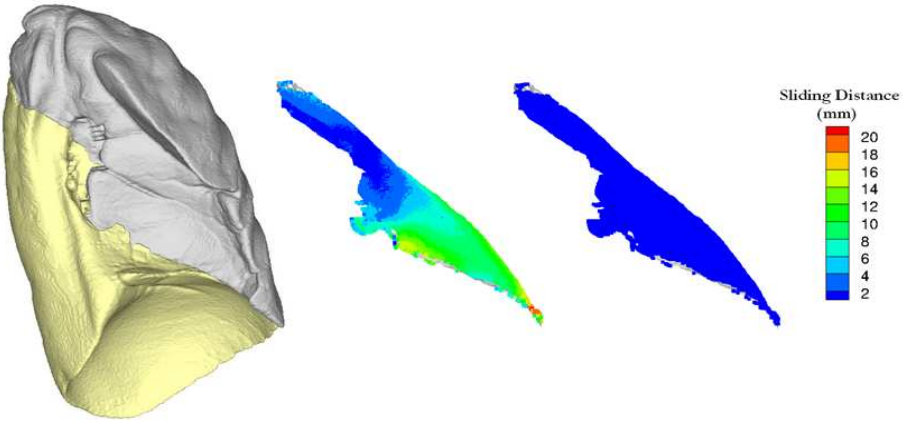
Figure 3 compares the estimated lobar sliding distances between two different registration methods. The whole-lung-based registration shows small sliding distance ($\leqslant$ 1 mm) because the transformation model enforces displacement field smoothness across the fissure, while the lobe-by-lobe registration method recovers the displacement field discontinuity along the fissure.

4 Discussion and Summary

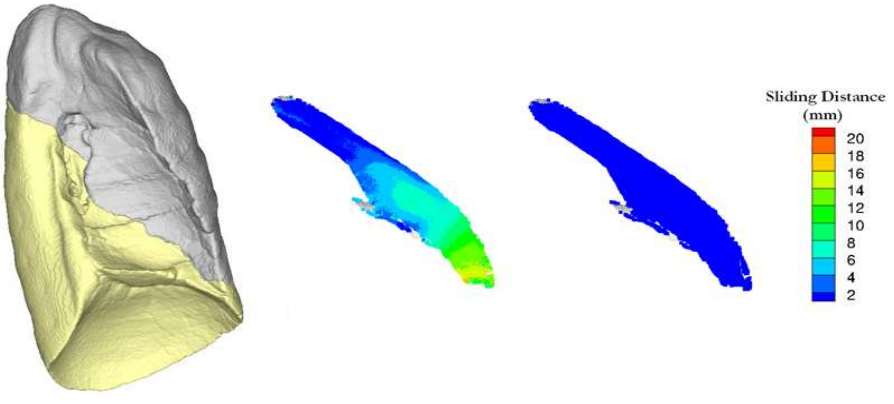
We proposed a method to estimate local lobe sliding using lobe-by-lobe lung-mass-preserving registration. In addition, we compared the displacement field, the landmark error, and the sliding distance between the lobe-by-lobe registration and the lung-by-lung registration for the left lungs of three normal human subjects. We used major vascular bifurcation as landmarks. Thus, there are not large numbers of landmarks near fissures. As seen in Figure 1, both methods yield similar matching results in the center regions of the lobes while a considerable difference is observed in the vicinity of fissure. Thus, as seen in Table 1 there is not a very significant decrease in the overall landmark error while there are significant regional differences. Moreover, the lung-by-lung registration is not able to capture the sliding between the lobes while the lobe-by-lobe registration shows the same superior-inferior gradient of sliding distance in all three cases. One possible explanation for this pattern is that the lungs contract and expand more at the diaphragm than at the apex and the LUL is more firmly anchored to the chest wall than LLL.

In addition to evaluating lobar sliding, the lobe-by-lobe registration may yield more physiologically meaningful assessments of regional lung function and mechanics. Registration transformation functions that do not explicitly model the lobar fissure are not able to capture lobar sliding and thus experience more registration errors near the fissure. These findings may have implications in using registration to estimate lung function (specific volume change and lung expansion) and for tracking lung tissue and lung nodules, across the respiratory cycle. These methods can be directly extended to respiratory-gated CT of the lung, where CT data is reconstructed at multiple points across the respiratory cycle.

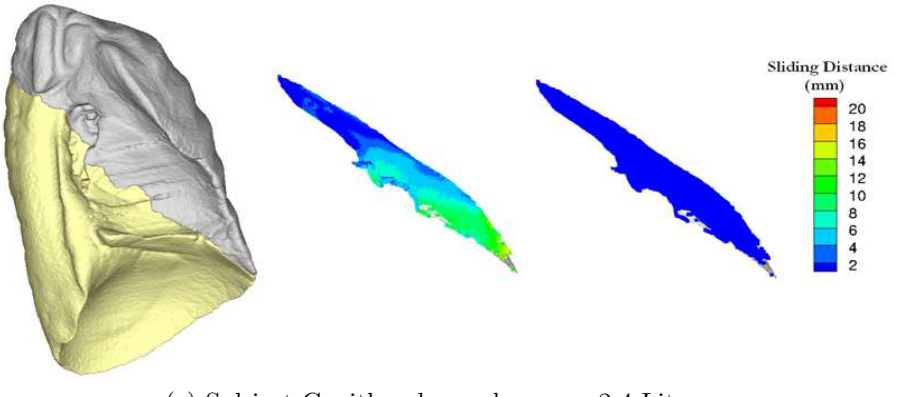
In conclusion, we have described a method to evaluate the local lobar sliding using a lobe-by-lobe lung-mass-preserving registration. Application of these



(a) Subject A with volume change = 3.2 Liters



(b) Subject B with volume change = 2.3 Liters



(c) Subject C with volume change = 2.4 Liters

Fig. 3. The color-coded sliding distance map overlays on the fissure surface. Left most column is the surface rendering of LUL (gray) and the LLL (gold); second column shows the sliding distance from the lobe-based registration; and right most column shows the sliding distance from the whole-lung registration.

methods may be useful for increasing our understanding of function and biomechanical behavior of the respiratory system.

Acknowledgments. The authors thank Ms. Keelin Murphy and Dr. Bram van Ginneken for providing the software iX for annotating landmarks, and Mr. Matthew Moehlmann and Mr. Luke Gabe for the data analysis. This work was supported in part by NIH grants HL079406, HL064368, EB005823, RR022421, EB004126, HL080285 and a University of Iowa CTSA NIH/NCRR grant 1UL1RR024979.

References

1. Hubmayr, R.D., Rodarte, J.R., Walters, B.J., Tonelli, F.M.: Regional ventilation during spontaneous breathing and mechanical ventilation in dogs. J. Appl. Physiol. 63(6), 2467–2475 (1987)
2. Christensen, G.E., Song, J.H., Lu, W., Naqa, I.E., Low, D.A.: Tracking lung tissue motion and expansion/compression with inverse consistent image registration and spirometry. Med. Physics 34(6), 2155–2165 (2007)
3. Guerrero, T., Sanders, K., Noyola-Martinez, J., Castillo, E., Zhang, Y., Tapia, R., Guerra, R., Borghero, Y., Komaki, R.: Quantification of regional ventilation from treatment planning CT. Int. J. Radiation Oncology Biol. Phys. 62(3), 630–634 (2005)
4. Reinhardt, J.M., Ding, K., Cao, K., Christensen, G.E., Hoffman, E.A., Bodas, S.V.: Registration-based estimates of local lung tissue expansion compared to xenon-CT measures of specific ventilation. Medical Image Analysis 12(6), 752–763 (2008)
5. Gee, J., Sundaram, T., Hasegawa, I., Uematsu, H., Hatabu, H.: Characterization of regional pulmonary mechanics from serial magnetic resonance imaging data. Acad. Radiol. 10, 1147–1152 (2003)
6. Ukil, S., Reinhardt, J.M.: Anatomy-guided lung lobar surface detection in X-ray CT images. IEEE Trans. Medical Imaging 28(2), 202 (2009)
7. Yin, Y., Hoffman, E.A., Lin, C.L.: Local tissue-weight-based nonrigid registration of lung images with application to regional ventilation, vol. 7262, SPIE, 72620C (2009)
8. Hoffman, E.A., Ritman, E.L.: Effect of body orientation on regional lung expansion in dog and sloth. J. Appl. Physiol. 59(2), 481–491 (1985)
9. Murphy, K., van Ginneken, B., Pluim, J.P.W., Klein, S., Staring, M.: Semi-automatic reference standard construction for quantitative evaluation of lung CT registration. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 1006–1013. Springer, Heidelberg (2008)

Evaluation of 4D-CT Lung Registration

Sven Kabus[1], Tobias Klinder[1,2], Keelin Murphy[3], Bram van Ginneken[3], Cristian Lorenz[1], and Josien P.W. Pluim[3]

[1] Philips Research Europe – Hamburg, Germany
sven.kabus@philips.com
[2] Institut für Informationsverarbeitung, Leibniz University of Hannover, Germany
[3] University Medical Center, Utrecht, The Netherlands

Abstract. Non-rigid registration accuracy assessment is typically performed by evaluating the target registration error at manually placed landmarks. For 4D-CT lung data, we compare two sets of landmark distributions: a smaller set primarily defined on vessel bifurcations as commonly described in the literature and a larger set being well-distributed throughout the lung volume. For six different registration schemes (three in-house schemes and three schemes frequently used by the community) the landmark error is evaluated and found to depend significantly on the distribution of the landmarks. In particular, lung regions near to the pleura show a target registration error three times larger than near-mediastinal regions. While the inter-method variability on the landmark positions is rather small, the methods show discriminating differences with respect to consistency and local volume change. In conclusion, both a well-distributed set of landmarks and a deformation vector field analysis are necessary for reliable non-rigid registration accuracy assessment.

1 Introduction

Image registration plays an indispensable role in medical application areas such as diagnosis, therapy planning, and follow-up assessment. Evaluating its accuracy has therefore become an issue of growing importance [1,2,3]. Contrary to rigid registration methods, validation of non-rigid methods remains a challenging task due to the absence of a gold standard [3]. Currently, this task is addressed by computing the target registration error (TRE) based on anatomical landmarks or on surfaces (e.g., contoured masses) in the images to be registered. The resulting TRE is widely accepted as an indicator for the method's accuracy. TRE-based evaluation for lung applications, however, suffers from low contrast in near-to-pleura regions, in particular if the scans are acquired dynamically as in 4D-CT. Consequently, evaluation is typically performed on a limited set of landmarks defined on major bifurcations of the vessel tree (cf., e.g., [4,1,5]), often grouped around the mediastinum and thus not covering the lungs entirely. Since registration accuracy in near-mediastinal regions may not be representative for, e.g., the lower lungs as here the motion amplitude is largest, it is therefore of particular interest how the choice of landmarks affects the estimated overall registration accuracy.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 747–754, 2009.

On the basis of a publicly available and carefully validated 4D-CT chest dataset (POPI-model [4]) we compare in this work the influence of two different landmark sets on the TRE: (a) a set consisting of 37 lung landmarks grouped around the mediastinum (obtained by the POPI-model), (b) a set of 100 landmarks being well-distributed throughout the lung volume (generated by a semi-automatic tool [6]). For the registration accuracy, six different registration algorithms are compared in terms of the TRE (per landmark set and per lung region) as well as the inter-method variability at the landmark positions. In addition, an analysis of the resulting deformation vector fields is included.

2 Methods

2.1 Datasets

The basis of the evaluation is a respiratory gated chest CT dataset (10 phases, 0.98x0.98x2.00mm^3 voxel size) provided by the 'POPI-model' [4]. Maximum diaphragm motion is measured as about 18mm (9mm) for the right (left) lung. In addition, the POPI-model provides a set of 41 landmarks annotated in each breathing phase. From these, 37 are located within the lungs and are employed for evaluation (*landmark reference set (LRS) A*). Moreover, an automatic tool [6] is used to create a set of 100 well-distributed lung landmarks in a reference phase. These landmarks are then semi-automatically propagated [6] to all other respiratory phases individually by two trained scientists (inter-observer difference of 1.0mm); their averaged annotations finally defines the second 4D set of landmark positions (*LRS B*). For a region-based analysis, each lung is divided into four volumetric regions such that the near-mediastinal parenchyma is enclosed in a sphere around the lungs' center of mass (regions I and II) and the remaining parenchyma of each lung is separated into an upper, central, and lower region. The dispersal of landmarks concerning the defined regions is shown in Fig. 1 for both landmark reference sets. While the landmarks from set A mainly belong to regions I and II, the landmarks from set B show a more regular distribution.

2.2 Registration Schemes

This work compares six fully automatic registration schemes, each computing deformation vector fields (DVFs) pointing from the end-inhale phase onto all other respiratory phases. Three methods are in-house, three other methods have either been downloaded or have been already applied to the data set. Five of the six methods are volumetric schemes whereas one is surface-based. Among the volumetric schemes, three are based on a parametric B-spline representation, while the two others are non-parametric.

Surface-based Registration (MBS). [7] After extracting lung surface and vessel tree a method of deformable surface models is applied on the generated

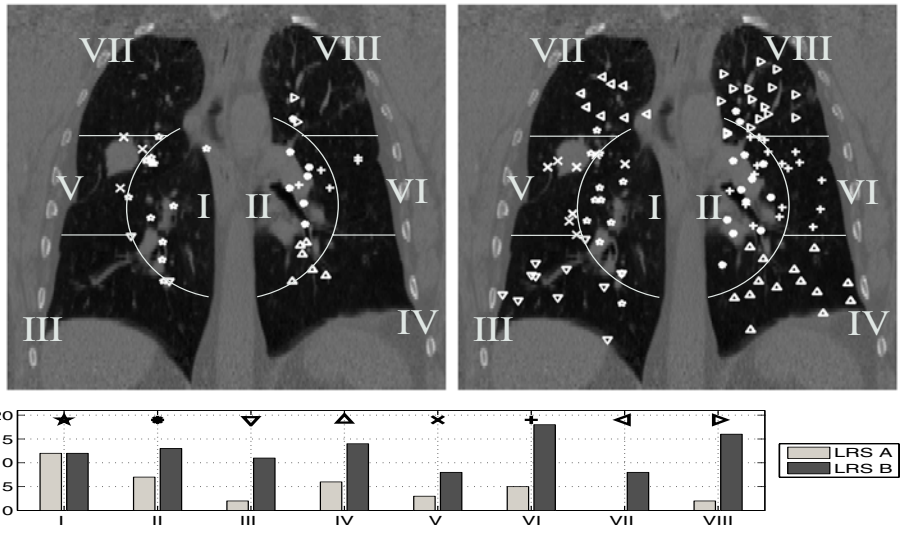


Fig. 1. Region definition and landmark positions shown in a coronal projection for the landmark reference set A (top left) and for the landmark reference set B (top right) together with the dispersal of landmarks among the regions (bottom). For improved visualization each region is given a unique landmark symbol.

Table 1. Property overview for the volumetric registration schemes. All schemes are used in a multi-level embedding.

Method	transformation class	cost function	regularizer	use of lung mask
POPI-par [8]	B-splines	SSD	–	no
elastix [9]	B-splines	NCC	–	yes
DROP [10]	B-splines	SAD	grid vector distances	yes
POPI-nonpar [5]	non-par.	SSD	Gaussian + linear elastic	no
FEIR [11]	non-par.	SSD	linear elastic	yes

iso-surface to propagate the mesh to all other phases by minimizing an energy term. A dense DVF is then obtained using a thin-plate-spline interpolation scheme.

Volumetric Registrations [5,8,9,10,11]. An overview of the employed volumetric schemes is given in Tab. 1, for further details we refer to the cited references. POPI-par is a modification based on a frequently used FFD-approach [12] whereas POPI-nonpar describes a modified Demon's approach [5].

MBS, elastix and FEIR are in-house methods and are applied with standard parameter settings. DROP is applied following the published instructions [10]. The computation for these methods was completed after 3–15 minutes per phase on standard PCs. The DVFs for POPI-par and POPI-nonpar are not computed by the authors themselves, these are already provided by the POPI-model.

3 Results

To ensure the correct use of the output of the different registration schemes, each DVF is applied to the reference phase to carefully check the correct alignment of the pleura with the corresponding phase.

Then, for all the six methods each set of landmarks annotated in the reference phase is transformed to the other phases. By comparing the transformed position with the respective annotated position, the target registration error (TRE) is obtained for each landmark, respiratory phase, and registration method. From this, we deduce both method-based and LRS-based mean TRE and standard deviation by collecting the TREs from all phases and landmarks, cf. Tab. 2. In addition, the mean of the upper 10% quantile (MU10) is calculated, which, compared to the maximum, is less dominated by outliers. An inspection of Tab. 2 reveals a clear decrease of the TRE for all methods with elastix performing slightly better and POPI-nonpar slightly worse than the others. Focusing on the inter-LRS differences, we further notice that the MU10 values show a larger relative difference than the mean values. The reason can be found in lung regions covered by LRS B but not by LRS A (cf. the landmark distribution displayed in Fig. 1) – on the one hand this refers to regions located in the apex of the lungs with a quite small TRE, on the other hand to regions in the lower lungs with a larger TRE (demonstrated in Fig. 2). Particularly in the lower right lung (region III) the mean-TRE (2.4–3.1mm) as well as the MU10-TRE (6.1–8.7mm) are higher by a factor three (cf. Tab. 3) probably caused by the larger motion amplitude in the right lung. However, since Fig. 1 reports only 2 landmarks in region III for LRS A, this inhomogeneous distribution of the TRE cannot be deduced from a landmark distribution grouped around the mediastinum.

As the TRE differs only slightly among the registration methods, we investigate the inter-method variability evaluated on the displaced landmark positions of both landmark sets. To this end, for each landmark position and each phase individually, the sum of eigenvalues of the covariance matrix is computed from the displacement vectors of all methods. For an LRS-based discrimination, its mean is taken over all landmarks and phases yielding a value of 0.31mm for set A and one of 0.47mm for set B, thus an increased inter-method variability

Table 2. TRE for each method and LRS (shown are mean, standard deviation, and MU10 in mm)

Method	Landmark reference set A	Landmark reference set B
without registration	3.68 ± 2.97 (9.6)	4.22 ± 3.20 (11.2)
MBS	1.11 ± 0.65 (2.5)	1.36 ± 1.13 (3.9)
POPI-par	1.07 ± 0.56 (2.2)	1.26 ± 1.17 (3.8)
elastix	0.96 ± 0.56 (2.1)	1.16 ± 0.97 (3.4)
DROP	0.98 ± 0.56 (2.2)	1.25 ± 1.07 (3.6)
POPI-nonpar	1.28 ± 0.42 (2.1)	1.41 ± 1.22 (4.1)
FEIR	1.05 ± 0.57 (2.2)	1.27 ± 1.00 (3.6)

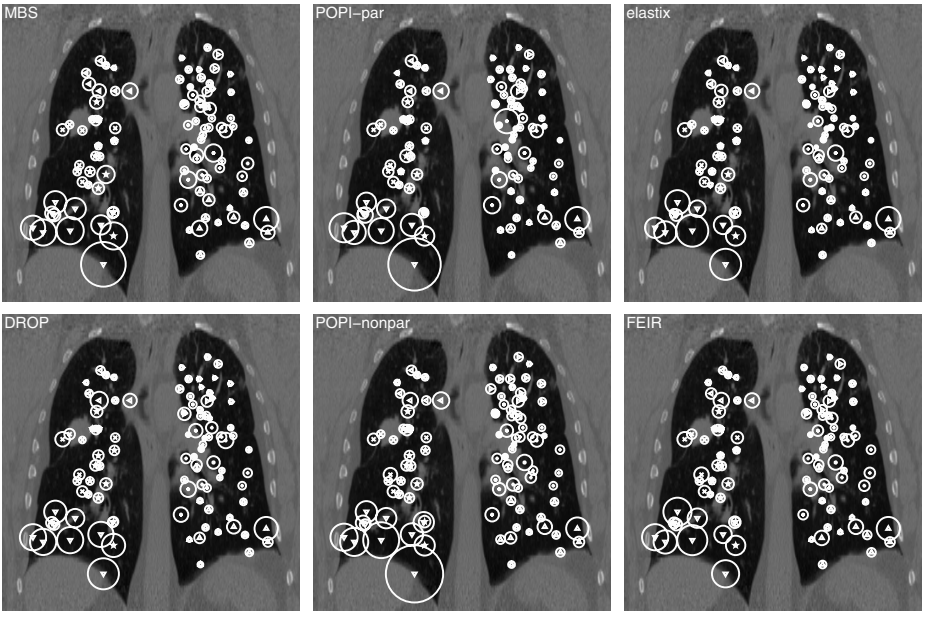


Fig. 2. TRE averaged over all respiratory phases for landmark reference set B shown proportional to the spheres' diameters

Table 3. TRE per lung region (cf. Fig. 1) for landmark reference set B (shown are mean, standard deviation, and MU10 in mm)

Method	Region I	Region II	Region III	Region IV
without reg.	5.28 ± 3.58 (11.5)	3.65 ± 2.00 (6.9)	7.31 ± 5.25 (15.7)	4.83 ± 2.77 (9.1)
MBS	1.44 ± 1.10 (3.7)	1.25 ± 0.73 (2.7)	2.69 ± 2.25 (7.4)	1.30 ± 0.87 (3.0)
POPI-par	1.21 ± 0.77 (2.6)	1.28 ± 0.96 (3.2)	2.71 ± 2.47 (7.9)	1.12 ± 0.86 (2.8)
elastix	1.17 ± 0.79 (2.8)	1.09 ± 0.69 (2.4)	2.37 ± 1.89 (6.1)	1.09 ± 0.80 (2.6)
DROP	1.33 ± 0.89 (3.1)	1.20 ± 0.75 (2.7)	2.49 ± 2.17 (7.4)	1.11 ± 0.80 (2.7)
POPI-nonpar	1.38 ± 0.77 (2.8)	1.22 ± 0.66 (2.5)	3.14 ± 2.60 (8.7)	1.25 ± 0.77 (2.6)
FEIR	1.21 ± 0.82 (2.9)	1.08 ± 0.66 (2.3)	2.48 ± 1.83 (6.2)	1.18 ± 0.86 (2.8)

Method	Region V	Region VI	Region VII	Region VIII
without reg.	4.64 ± 3.12 (10.4)	3.64 ± 1.94 (6.8)	3.58 ± 2.03 (7.1)	1.99 ± 1.15 (4.2)
MBS	1.24 ± 0.68 (2.7)	0.90 ± 0.46 (1.9)	1.12 ± 0.57 (2.3)	0.89 ± 0.47 (1.9)
POPI-par	1.15 ± 0.72 (2.7)	0.92 ± 0.50 (2.0)	1.11 ± 0.55 (2.2)	0.73 ± 0.39 (1.5)
elastix	1.27 ± 0.65 (2.6)	1.01 ± 0.57 (2.2)	1.14 ± 0.66 (2.5)	0.82 ± 0.45 (1.8)
DROP	1.19 ± 0.62 (2.4)	1.12 ± 0.62 (2.5)	1.30 ± 0.60 (2.4)	0.90 ± 0.41 (1.7)
POPI-nonpar	1.29 ± 0.73 (2.9)	1.10 ± 0.53 (2.2)	1.28 ± 0.66 (2.6)	1.01 ± 0.44 (1.8)
FEIR	1.41 ± 0.83 (3.1)	1.10 ± 0.62 (2.5)	1.17 ± 0.64 (2.4)	0.86 ± 0.44 (1.7)

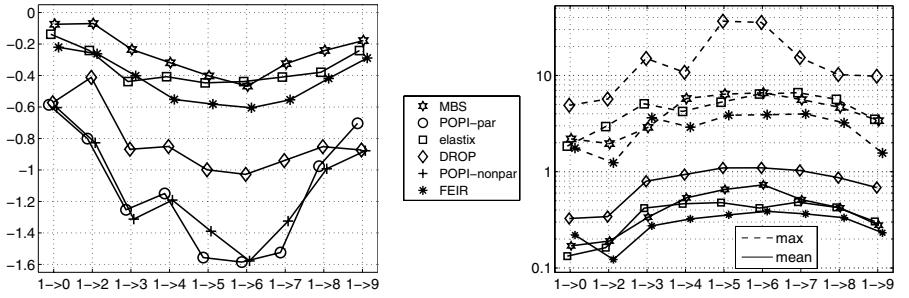


Fig. 3. Left: Minimum volume change (obtained from $\det(\mathrm{Jac}(\mathrm{DVF})) - 1$) over all lung voxels shown for each method and phase transition (values smaller than -1 indicate folding of the DVF); right: Mean and maximum consistency error (in mm) over all lung voxels shown for each phase transition

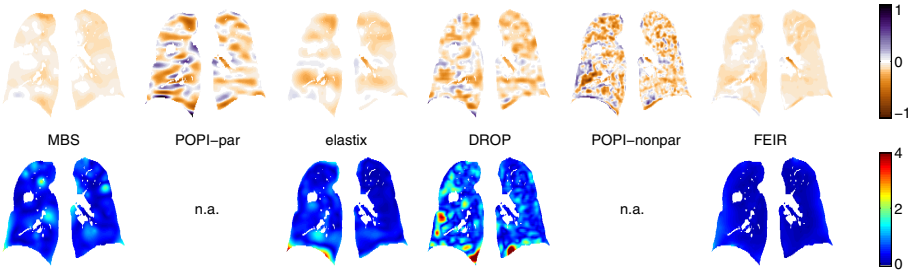


Fig. 4. For an exemplary coronal slice (same as in Fig. 1), local volume change (top) and local consistency error (bottom) are displayed after registration from end-inhale onto end-exhale. Note that for the local volume change maps blue (brown) colors indicate expansion (contraction).

on the well-distributed landmark set. Discriminating on a region-by-region basis determines a value of 1.81mm for region III and a range of 0.24–0.36mm for the other regions, thus a high variability for regions with a large TRE and vice versa. Finally, from a computation of the Mahalanobis distance, elastix and FEIR are identified to deviate less from the other methods (mean values of 2.21 and 2.09, respectively) whereas POPI-nonpar shows the largest deviation (6.66).

However, evaluation so far is restricted to the sparse sets of distinct landmark positions. To further investigate the inter-method variability and also to achieve greater insight into the methods' properties an analysis of the DVFs is performed.

At first, we compute the determinant of the Jacobian of the DVF to measure the local change in volume which allows for detecting non-bijectively mapped regions (i.e., 'folding' occurs). Since each DVF is defined in the end-inhale domain we expect a contraction in general. An inspection of the minimum values (within the lungs only) revealed large differences: whereas MBS, elastix and FEIR show relatively little local contraction values, DROP and the two POPI-methods result in heterogeneous contraction-expansion-patterns (cf. Fig. 4, top) with severe foldings of the POPI-methods for phases 3 to 7 (cf. Fig. 3, left).

Secondly, for registration methods with both forward and backward DVFs available, we compute a local consistency measure. The consistency is determined by composing the forward and the backward DVFs and measuring the amplitude of the geometrical discrepancy at each position. Since none of the methods under consideration is consistent by definition, measuring the consistency error is a suitable indicator of how independent the registration result is from the image input order. Here, the smallest error is achieved by FEIR, the largest by DROP (cf. Fig. 3, right, and Fig. 4, bottom). Neither POPI-method is evaluated since backward DVFs were required.

4 Discussion and Conclusion

On the basis of a publicly available 4D-CT chest dataset we have investigated the impact of two different landmark distributions on non-rigid registration accuracy assessment: the first set of landmarks consists of 37 lung annotations grouped around the mediastinum as commonly described in the literature, the second set provides 100 annotations being well-distributed throughout the lung volume. While the mean target registration error on both landmark sets differs only slightly, a region-based analysis reveals smaller errors in apical regions but also a significantly higher error in the lower right lung and therefore a dependence of the target registration error on the distribution of landmarks. This dependence is observed for each registration scheme chosen from a collection of six popular methods including surface- and volume-based as well as parametric and non-parametric methods. Clearly, validation on further patient datasets is required, however, since generally the lower lungs show the largest motion amplitude, we believe that our result is of general validity and that a well-distributed set of landmarks is a necessity for reliable registration accuracy assessment.

Focusing on the inter-method variability we observed similar displacement vectors at the landmark positions. The methods can, however, be discriminated by a closer analysis of the resulting deformation vector fields, e.g., regarding the local change in volume. This parameter is clinically relevant since it is directly related to local lung ventilation estimation and has impact on lung diagnosis as well as on radiotherapy planning. Inspecting the local volume change discloses for three of the six methods several regions where the vector field is near to or already suffers from folding. Moreover, the methods differ in showing individual contraction-expansion-patterns. With the lung at end-inhale as the reference position in mind, an overall contraction – rather than a mixture of contracting and expanding regions – seems to be desirable from a physiological point of view. Another indication for inter-method variability is obtained from a local consistency analysis. Although non-consistent by definition, three out of four methods result in nearly consistent deformation vector fields and are thus independent of the image input order. Since similar target registration errors does not mean that different registration methods give similar results overall, an analysis of the deformation vector field is necessary to evaluate non-rigid registration accuracy.

Acknowledgments. We thank Heike Ruppertshofen for carrying out the landmark propagation as one observer. The POPI-dataset was acquired from the Léon Bérard Cancer Center, Lyon, France.

References

1. Brock, K.K.: On Behalf of the Deformable Registration Accuracy Consortium: A Multi-Institution Deformable Registration Accuracy Study. In: I. J. Radiation Oncology, ASTRO 2007, vol. 69, p. S44 (2007)
2. Vik, T., Kabus, S., von Berg, J., Ens, K., Dries, S., Klinder, T., Lorenz, C.: Validation and comparison of registration methods for free-breathing 4D lung CT. In: Proc. of SPIE Medical Imaging, vol. 6914, 69142P–1–69142P–10 (2008)
3. Christensen, G.E., Geng, X., Kuhl, J.G., Bruss, J., Grabowski, T.J., Pirwani, I.A., Vannier, M.W., Allen, J.S., Damasio, H.: Introduction to the non-rigid image registration evaluation project (NIREP). In: Pluim, J.P.W., Likar, B., Gerritsen, F.A. (eds.) WBIR 2006. LNCS, vol. 4057, pp. 128–135. Springer, Heidelberg (2006)
4. Vandemeulebroucke, J., Sarrut, D., Clarysse, P.: The POPI-model, a point-validated pixel-based breathing thorax model. In: Proc. of ICCR (2007)
5. Sarrut, D., Boldea, V., Miguet, S., Ginestet, C.: Simulation of 4D CT images from deformable registration between inhale and exhale breath-hold CT. Med. Phys. 33(3), 605–617 (2006)
6. Murphy, K., van Ginneken, B., Pluim, J.P.W., Klein, S., Staring, M.: Semi-automatic reference standard construction for quantitative evaluation of lung CT registration. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 1006–1013. Springer, Heidelberg (2008)
7. von Berg, J., Barschdorf, H., Blaffert, T., Kabus, S., Lorenz, C.: Surface based cardiac and respiratory motion extraction for pulmonary structures from multiphase CT. In: Proc. of SPIE Medical Imaging, vol. 6511, 65110Y (2007)
8. Delhay, B., Clarysse, P., Pera, C., Magnin, I.E.: A spatio-temporal deformation model for dense motion estimation in periodic cardiac image sequences. In: Proc. of the workshop From Statistical Atlases to Personalized Models: Understanding Complex Diseases in Populations and Individuals, MICCAI, pp. 87–90 (2006)
9. Klein, S., Staring, M., Pluim, J.P.W.: Evaluation of optimization methods for non-rigid medical image registration using mutual information and B-splines. IEEE Trans. Image Proc. 16(12), 2879–2890 (2007)
10. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through MRFs and efficient linear programming. Med. Image Analysis 12(6), 731–741 (2008)
11. Kabus, S., von Berg, J., Yamamoto, T., Opfer, R., Keall, P.J.: Lung ventilation estimation based on 4D-CT imaging. In: Proc. of the first international workshop on pulmonary image analysis, MICCAI, pp. 73–82 (2008)
12. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Trans. Med. Imaging 18(8), 712–721 (1999)

Slipping Objects in Image Registration: Improved Motion Field Estimation with Direction-Dependent Regularization

Alexander Schmidt-Richberg, Jan Ehrhardt, Rene Werner, and Heinz Handels

Department of Medical Informatics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
a.schmidt-richberg@uke.uni-hamburg.de

Abstract. The computation of accurate motion fields is a crucial aspect in 4D medical imaging. It is usually done using a non-linear registration without further modeling of physiological motion properties. However, a globally homogeneous smoothing (regularization) of the motion field during the registration process can contradict the characteristics of motion dynamics. This is particularly the case when two organs slip along each other which leads to discontinuities in the motion field. In this paper, we present a diffusion-based model for incorporating physiological knowledge in image registration. By decoupling normal- and tangential-directed smoothing, we are able to estimate slipping motion at the organ borders while ensuring smooth motion fields in the inside and preventing gaps to arise in the field. We evaluate our model focusing on the estimation of respiratory lung motion. By accounting for the discontinuous motion of visceral and parietal pleurae, we are able to show a significant increase of registration accuracy with respect to the target registration error (TRE).

1 Introduction

The accessibility of accurate motion fields is a precondition for many applications in medical imaging, e.g. for lung motion quantification in radiation therapy [1], wall movement analysis of the heart [2] or automatic contour propagation [3]. Motion estimation is usually done on the basis of 4D images by computing the displacements between time frames of the data sets using non-linear registration approaches.

While smoothing mechanisms are required during the registration process, they can contradict the physiology of the organ motion. This is especially the case when two objects slip along each other, which can be observed for example in the case of lung- [4] or liver motion [5].

Arising discontinuities at object boundaries are addressed by several approaches: In [4] finite element methods are used in order to simulate the physiology of respiration dynamics, point- and surface-based registration approaches are used in [6]. While these allow to explicitly or implicitly incorporate boundary conditions, inner-organ information like bronchial or vessel trees are dismissed

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 755–762, 2009.

which results in an inaccurate registration of those structures. In [5,7] the problem of slipping organs is dressed implicitly by masking the background in order to prevent it from affecting the force calculation. However, this limits motion estimation to the object and provides no information about background motion.

In this paper, we present a novel smoothing mechanism that enables us to explicitly handle slipping motion at object boundaries. While other approaches use segmentations in order to model certain properties of objects (e.g. rigidity [8] or varying elasticities [9]) none addresses this specific physiological behaviour. Our model is closely related to the diffusive regularisation approach [10]. By decoupling normal- and tangential-directed smoothing we are able to estimate discontinuities at the object boundaries while maintaining smooth inner-object motion and preventing gaps at the borders.

The presented approach is not limited to a specific application. However, we apply it in the context of radiotherapy for the treatment of thoracic tumors, where the computation of accurate motion fields of the lung is a crucial part. In-depth knowledge about lung motion is needed to explicitly account for the respiration-driven movement of tumor and organs at risk during the treatment, for example in order to determine appropriate safety margins or dose accumulations [1]. From the perspective of physiology, motion estimation is especially challenging because visceral and parietal pleurae are slipping along each other during breathing, thus creating discontinuities in lung and chest wall motion. Transferred to the registration problem, this behavior conflicts with common regularisation models which favor smooth motion fields. As a result, registration errors arise in particular near the lung borders [7].

This paper is organized as follows. We start introducing our model in section 2. In section 3 a detailed evaluation and discussion is given for the estimation of lung motion fields. We summarize our approach in section 4.

2 Methods

When applying non-linear registration methods for the estimation of organ motion, usually no a-priori knowledge about the physiological process is taken into account. Treating all regions in the image as one homogeneous object often results in questionable motion estimates.

In this paper, we present a model for a normal-directed regularization. We incorporate knowledge about the physiology of organ motion by allowing object and background to slide along each other (see Fig. 1). In the process we assume a given segmentation of the object in one frame.

We proceed by presenting an automatic preprocessing step to refine the segmentations, what turns out to be useful in order to improve accuracy and stability of subsequent steps (Sect. 2.1). In Sect. 2.2 we briefly summarize the basics of the diffusive registration. We then introduce our model for the novel regularisation approach in Sect. 2.3.

2.1 Preprocessing of Segmentations

In order to determine object boundaries we use segmentations of the object, i.e. of the lung in our particular case. These segmentations were generated by a semi-automatic procedure, consisting of a volume growing followed by a manual slice-by-slice correction of the boundaries by a clinical expert. Due to this generation process and reconstruction artifacts in the CT data (comp. [11]), the segmentations often show inaccuracies and a non-smoothness, especially in z-direction.

These shortcomings are particularly impedimental to the discrete calculation of the surface normals (comp. Sect. 2.3). Therefore, we apply a level set based refinement of the segmentations.

Let Γ be the segmented object in the reference image and $\widetilde{\phi} : \Omega \mapsto \mathbb{R}$ the corresponding level set function, with $\Omega \subset \mathbb{R}^3$ denoting the image domain. We calculate an improved segmentation $\phi(\boldsymbol{x})$, $\boldsymbol{x} \in \Omega$ by minimizing the following energy functional:

$$\mathcal{J}_{Seg}[\phi] := \mathcal{I}[\phi] + \mathcal{E}[R;\phi] + \mathcal{A}[\widetilde{\phi};\phi]$$

The internal energy $\mathcal{I}$ provides a smoothing of the boundary as defined in [12]. We chose a region based external energy $\mathcal{E}$ as proposed in [3].

The third energy term

$$\mathcal{A}[\widetilde{\phi};\phi] := \frac{1}{2}\int_{\Omega} H_\alpha(\phi(\boldsymbol{x}))\|\widetilde{\phi}(\boldsymbol{x}) - \phi(\boldsymbol{x})\|^2 \, d\boldsymbol{x} \ ,$$

where H_α denotes a differentiable approximation of the Heaviside function, prevents the segmentation from moving too far away from the manual segmentation. As a result, we obtain a segmentation with increased smoothness and accuracy.

2.2 Diffusive Registration

Let $R, T : \Omega \mapsto \mathbb{R}$ be two 3D Images (i.e. timeframes) of a 4D data set, called reference image $R(\boldsymbol{x})$ and target image $T(\boldsymbol{x})$ with the image domain $\Omega \subset \mathbb{R}^3$. The goal is to find a motion field $\boldsymbol{u} : \Omega \mapsto \mathbb{R}^3$ that minimizes the energy functional

$$\mathcal{J}_{Reg}[\boldsymbol{u}] := \mathcal{D}[R, T; \boldsymbol{u}] \ + \ \mathcal{S}[\boldsymbol{u}] \tag{1}$$

The distance measure $\mathcal{D}$ is used to measure the similarity between the reference image and the warped target image $T(\boldsymbol{x} - \boldsymbol{u}(\boldsymbol{x}))$ whereas the smoothing term $\mathcal{S}$ provides smooth motion fields.

A common choice for $\mathcal{S}$ is the diffusive registration [10]

$$\mathcal{S}^{Diff}[\boldsymbol{u}] := \frac{1}{2}\sum_{l=1}^{3}\int_{\Omega} \|\nabla u_l(\boldsymbol{x})\|^2 \, d\boldsymbol{x} \ , \tag{2}$$

where u_l is the l-th component of $\boldsymbol{u}$. A smoothing is archieved by penalizing large gradients in the vector field. Providing results similar to those of the elastic regularisation, this term is much more efficient with respect to computation time [7].

For the minimization of the energy functional (1) an iterative approach is employed. According to the calculus of variations, a solver of (1) must fulfill the condition $\partial\mathcal{J}/\partial\boldsymbol{u} = 0$. From the associated Euler-Lagrangian equation the following iterative update scheme is derived:

$$\boldsymbol{u}^{(k+1)} = \boldsymbol{u}^{(k)} + \tau\boldsymbol{c}_{\mathcal{S}}(\boldsymbol{u}^{(k)}) + \tau\boldsymbol{f}_{\mathcal{D}}(\boldsymbol{u}^{(k)}) \ .$$

The force term $\boldsymbol{f}_{\mathcal{D}}(\boldsymbol{u})$ corresponds to the specific distance measure $\mathcal{D}$. In this paper, we use

$$\boldsymbol{f}_{\mathcal{D}}(\boldsymbol{u}) := \frac{R(\boldsymbol{x}) - T(\boldsymbol{x} - \boldsymbol{u}(\boldsymbol{x}))}{\|\nabla R(\boldsymbol{x})\|^2 + \kappa^2}\nabla R(\boldsymbol{x}) \ ,$$

where κ denotes a normalizer which is used in order to account for image contrast. This force term is closely related to Thirion's demons, the corresponding distance measure can be found in [10].

The correction term $\boldsymbol{c}_{\mathcal{S}}(\boldsymbol{u})$ is used to smooth the field according to the regularizer $\mathcal{S}$. For (2) we find $\partial\mathcal{S}/\partial\boldsymbol{u} = \Delta\boldsymbol{u}$ what leads to $\boldsymbol{c}_{\mathcal{S}}(\boldsymbol{u}) = \Delta\boldsymbol{u}$.

2.3 Model for a Directional-Dependent Regularisation

In this work, we extend the diffusive regularisation by restricting inter-object smoothing to the normal direction. Based on a given segmentation $\phi(\boldsymbol{x})$ of an anatomical object, the goal is to incorporate knowledge about the physiology of its motion by allowing object and background to slip along each other. From a technical point of view, we want to allow discontinuities between the movement of object and background in tangential direction, while maintaining smoothness in normal direction (see Fig. 1).

Let $\boldsymbol{n}(\boldsymbol{x}) = \nabla\phi(\boldsymbol{x})/\|\nabla\phi(\boldsymbol{x})\|$ be the normal of the segmentation at a point $\boldsymbol{x}$. We proceed by splitting the motion in two parts: the normal-directed part $\boldsymbol{u}^{\perp}(\boldsymbol{x}) = \langle\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{n}(\boldsymbol{x})\rangle\boldsymbol{n}(\boldsymbol{x})$ and the tangential-directed part $\boldsymbol{u}^{\parallel}(\boldsymbol{x}) = \boldsymbol{u}(\boldsymbol{x}) - \langle\boldsymbol{u}(\boldsymbol{x}), \boldsymbol{n}(\boldsymbol{x})\rangle\boldsymbol{n}(\boldsymbol{x})$. We can then rewrite eq. (2) as

$$\mathcal{S}^{Diff}[\boldsymbol{u}] = \frac{1}{2}\sum_{l=1}^{3}\int_{\Omega}\|\nabla(u_l^{\perp} + u_l^{\parallel})\|^2 \, d\boldsymbol{x} = \frac{1}{2}\sum_{l=1}^{3}\int_{\Omega}\|\nabla u_l^{\perp}\|^2 + \|\nabla u_l^{\parallel}\|^2 \, d\boldsymbol{x} \ , \quad (3)$$

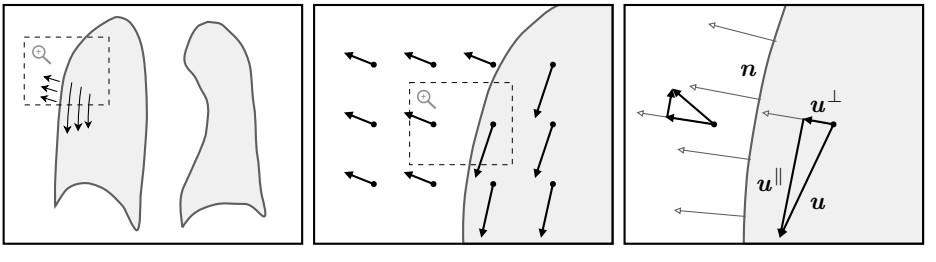


Fig. 1. In the center image the motion field along the border of the lung (left) is visualized. This is not smooth with respect to the diffusive regularization (2). By decoupling $\boldsymbol{u}^{\perp}$ and $\boldsymbol{u}^{\parallel}$ we can demand the field to be smooth only in normal direction (right).

whereat we assume $\boldsymbol{n}$ to be fixed, leading to $\sum_{l=1}^{3} \langle \nabla u_l^{\perp}, \nabla u_l^{\parallel} \rangle = 0$. This assumption is feasible because the surfaces are smooth with respect to image spacing.

Based on (3), we define our new model in two steps. First, according to our assumptions, we want to achieve a comprehensive (i.e. inter-object) smoothing in normal direction while smoothing object and background separately in tangential direction. Thus, we split the domain of the energy terms: as before we define $\|\nabla u_l^{\perp}\|^2$ on the whole image domain Ω, but we restrict $\|\nabla u_l^{\parallel}\|^2$ to the inside resp. outside of the object Γ, using Neumann boundary conditions. This leads to an interim formulation for a direction-dependent regularization (DDR)

$$\mathcal{S}^{DDR'}[\boldsymbol{u}] := \frac{1}{2} \sum_{l=1}^{3} \left(\int_{\Omega} \|\nabla u_l^{\perp}\|^2 \, d\boldsymbol{x} \; + \; \int_{\Gamma} \|\nabla u_l^{\parallel}\|^2 \, d\boldsymbol{x} \; + \; \int_{\Omega/\Gamma} \|\nabla u_l^{\parallel}\|^2 \, d\boldsymbol{x} \right) . \tag{4}$$

In a second step, we include a weighting between the proposed regularization (4) and the common diffusive term (2) in order to restrict the calculation of (4) to the region close to the object borders. This is done because normals are only defined within this region. Moreover, it entails a computational benefit and allows us to admit a tangential smoothing of a certain amount. Following [9] we use the Dirac-shaped weighting function

$$\delta(\phi(\boldsymbol{x})) = 1 - \frac{1}{1 + \alpha c \exp^{-\alpha \phi(\boldsymbol{x})^2}} \tag{5}$$

in order to determine the object borders. The influence of the parameters α and c on δ is depicted in [9]. This leads to the final energy term

$$\begin{aligned} \mathcal{S}^{DDR}[\boldsymbol{u}] := \frac{1}{2} \sum_{l=1}^{3} \Bigg(& \int_{\Omega} \delta(\phi) \; \|\nabla u_l^{\perp}\|^2 \; + (1 - \delta(\phi)) \; \|\nabla u_l\|^2 \, d\boldsymbol{x} \\ & + \int_{\Gamma} \delta(\phi) \; \|\nabla u_l^{\parallel}\|^2 \, d\boldsymbol{x} \; + \; \int_{\Omega/\Gamma} \delta(\phi) \; \|\nabla u_l^{\parallel}\|^2 \, d\boldsymbol{x} \Bigg) . \end{aligned} \tag{6}$$

As in Sect. 2.2, we apply a variational framework in order to minimize the energy functional (1). Regarding the Euler-Lagrangian equation of (6) we find

$$\boldsymbol{c}_{\mathcal{S}}(\boldsymbol{u}) = \nabla \delta \nabla \boldsymbol{u}^{\perp} \; + \; \nabla (1 - \delta) \nabla \boldsymbol{u} \; + \; \nabla_{\Gamma} \delta \nabla_{\Gamma} \boldsymbol{u}^{\parallel} \, . \tag{7}$$

With ∇_{Γ} we denote the gradient being calculated only inside the object, using Neumann boundary conditions. In the background, $\nabla_{\Omega/\Gamma}$ is used respectively. N.B.: we assume the normals $\boldsymbol{n}$ to be independent of the position $\boldsymbol{x}$ for the calculation of this derivative. The inclusion of the weighting function δ in (6) leads to terms related to an anisotropic diffusion which are implemented efficiently according to [13].

3 Results

The evaluation is based on 4D CT data sets of 12 lung cancer patients acquired during free breathing. Data sets are reconstructed using an optical flow based

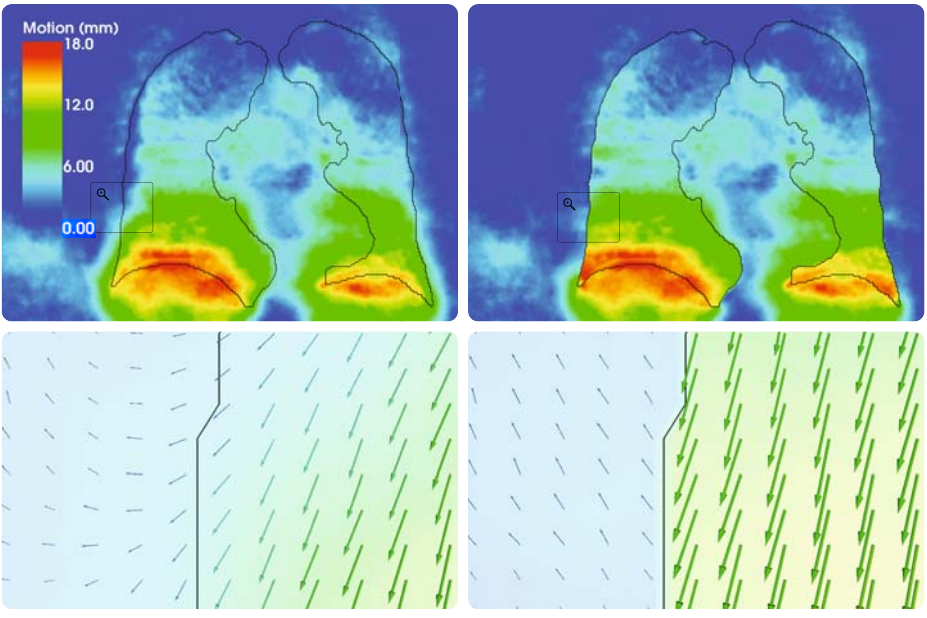


Fig. 2. Top: Motion magnitudes after a registration with diffusive (left, (2)) and directional-dependent regularisation (right, (6)). **Bottom:** The estimated motion field close to the lung border for the diffusive (left) and the directional-dependent (right) approach. For further explanation please refer to the text.

method [11]. Spatial resolution is $0.98 \times 0.98 \times 1.5$ mm. Each data set consists of 3D CT images at 10 to 14 different breathing phases of which we chose end-inspiration (EI) as reference image R and end-expiration (EE) as target image T. The segmentations are obtained as described in Sect. 2.1.

In Fig. 2, a qualitative comparison between the registration with the standard diffusive regularizer (2) and the direction-dependent regularizer (6) is given. In order to improve results as well as computational efficiency, a multi-resolution strategy was applied on both approaches. While the motion fields in the inside of the lung are almost identical, considerable differences occur near to the lung borders. In this area, the motion estimated by the directional-dependent approach satisfies the expectation derived from respiration physiology (comp. Fig. 1) and is not impaired by the marginal movement of the background as it is the case with the diffusive registration.

For a quantitative analysis, a medical expert identified inner lung landmarks (prominent bifurcations of the bronchial tree and the vessel tree) in the CT data at EI and EE. Landmarks are differentiated into landmarks located in the middle of the lung (20 landmarks each lung), near the lung borders (15 landmarks), and near the tumor (10 each lung tumor).

Registration accuracy is quantified as target registration error (TRE), i.e. the mean Euclidean distance between reference landmarks and target landmarks a) before registration, b) after a diffusive registration, c) after a masked

Table 1. Target registration errors (TRE) and variations in mm. The last row lists the statistical significance of the improvement of direction-dependent vs. diffusive regularisation, using levels of $p < 0.05$ (+), $p < 0.01$ (++) and $p < 0.001$ (+++).

Pat.	*w/o Reg.*	*Diffusive*	*Masked*	*Direction-Dep.*	*p-value*	*Significance*
01	4.25 ± 6.5	1.08 ± 0.7	1.06 ± 0.6	1.04 ± 0.4	0.422	=
02	6.26 ± 22.7	1.73 ± 1.9	1.21 ± 0.4	1.21 ± 0.4	< 0.001	+++
03	5.45 ± 7.4	1.73 ± 3.0	1.65 ± 2.8	1.48 ± 1.3	0.011	+
04	6.20 ± 4.6	1.74 ± 1.7	1.39 ± 0.7	1.44 ± 0.9	0.003	++
05	6.79 ± 14.0	1.80 ± 2.3	1.50 ± 1.4	1.55 ± 1.4	0.012	+
06	6.44 ± 9.5	1.67 ± 1.3	1.36 ± 0.6	1.43 ± 1.0	0.011	+
07	4.31 ± 9.8	1.41 ± 1.3	1.35 ± 0.6	1.35 ± 0.7	0.025	+
08	10.76 ± 60.8	2.48 ± 10.2	1.43 ± 1.6	1.49 ± 2.2	< 0.001	+++
09	6.40 ± 46.0	2.80 ± 11.5	2.52 ± 10.8	2.40 ± 7.2	0.002	++
10	6.06 ± 25.4	1.37 ± 2.5	1.10 ± 0.4	1.19 ± 0.9	0.063	=
11	7.98 ± 28.0	2.07 ± 6.9	1.69 ± 2.8	1.71 ± 3.6	0.007	++
12	8.31 ± 40.9	2.14 ± 5.4	1.64 ± 2.8	1.52 ± 1.6	< 0.001	+++

diffusive registration and d) after the directional-dependent registration. The masked registration is done following for example [5,7] by defining $\boldsymbol{f}_{\mathcal{D}}(\boldsymbol{x}) := 0$ for all $\boldsymbol{x} \in \Omega/\Gamma$. In order to quantify the improvement of the directional-dependent approach with relation to the standard diffusive registration, we further perform a two-sample t-test following [14]. The results are given in Table 1.

The presented directional-dependent model outperforms the diffusive approach with respect to the TRE for all examined patients. In 10 out of 12 cases, the results are statistically significant. As expected, the improvement is considerable especially for landmarks near the lung border. Neither the directional-dependent nor the masked registration proves to be superior with regard to registration accuracy. However, the directional-dependent approach allows a closed mathematical formulation and provides a motion estimation for the whole image domain while the masked registration limits calculation to the lung.

4 Discussion and Conclusion

We presented a new registration approach that enables us to incorporate physiological knowledge into the smoothing mechanism. In order to model the slipping motion of objects, regularization procedure was decoupled into normal and tangential direction. While smoothing object and background separately in normal direction, tangential-directed smoothing was performed comprehensively. In addition, we presented an approach for an automatic refinement of the segmentations which is substantial to warrant smooth normals.

The parameters c and α in (5) can be determined generally and are not very sensitive to the specific application. In addition, no further parameters occur compared to the standard approach.

We validated the approach for motion field estimation of the lung on 12 clinical data sets. The presented directional-dependent regularization significantly

outperformed the common diffusive approach with respect to the TRE and provided results similar to those of the masked registration. However, with the directional-dependent registration we introduced a model for the estimation of the whole image domain, which also includes neighboring structures.

References

1. Keall, P.J., Mageras, G.S., Balter, J.M., et al.: The management of respiratory motion in radiation oncology report of AAPM Task Group 76. Med. Phys. 33(10), 3874–3900 (2006)
2. Chandrashekara, R., Mohiaddin, R.H., Rueckert, D.: Cardiac motion tracking in tagged MR images using a 4D B-spline motion model and nonrigid image registration. In: Proc. ISBI, vol. 1, pp. 468–471 (2004)
3. Schmidt-Richberg, A., Ehrhardt, J., Handels, H.: Integrated Segmentation and Non-Linear Registration for Organ Segmentation and Motion Field Estimation in 4D CT Data. Methods Inf. Med. (in press, 2009)
4. Werner, R., Ehrhardt, J., Schmidt, R., Handels, H.: Patient-Specific Finite Element Modeling of Respiratory Lung Motion using 4D CT Image Data. Med. Phys. 35(5), 1500–1511 (2009)
5. von Siebenthal, M., Szekely, G., Gamper, U., Boesiger, P.: 4D MR imaging of respiratory organ motion and its variability. Phys. Med. Biol. 52(6), 1547–1564 (2007)
6. von Berg, J., Barschdorf, H., Blaffert, T., Kabus, S.: Surface based cardiac and respiratory motion extraction for pulmonary structures from multi-phase CT. In: Proc SPIE, vol. 6511, p. 65110Y (2007)
7. Werner, R., Ehrhardt, J., Schmidt-Richberg, A., Handels, H.: Validation and comparison of a biophysical modeling approach and non-linear registration for estimation of lung motion fields in thoracic 4D CT data. In: Proc SPIE (2009)
8. Pitiot, A., Guimond, A.: Geometrical regularization of displacement fields for histological image registration. Med. Image Anal. 1(12), 16–25 (2008)
9. Kabus, S.: Multiple-Material Variational Image Registration. PhD thesis, Institut für Mathematik, Universität zu Lübeck (2006)
10. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, Oxford (2004)
11. Ehrhardt, J., Werner, R., Säring, D., Lu, W., Low, D.A., Handels, H.: An optical flow based method for improved reconstruction of 4D CT data sets acquired during free breathing. Med. Phys. 34(2), 711–721 (2007)
12. Chan, T.F., Vese, L.A.: Active contours without edges. IEEE Trans. Image Process 10(2), 266–277 (2001)
13. Weickert, J., Romeny, B., Viergever, M.: Efficient and reliable schemes for nonlinear diffusion filtering. IEEE Trans Image Process 7(3), 398–410 (1998)
14. Sarrut, D., Delhay, S., Villard, P., et al.: A comparison framework for breathing motion estimation methods from 4-D imaging. IEEE Trans. Med. Imag. 26(12), 1636–1648 (2007)

Multi-modal Registration Based Ultrasound Mosaicing

Oliver Kutter[1], Wolfgang Wein[2], and Nassir Navab[1]

[1] Computer Aided Medical Procedures, TUM, Munich, Germany
{kutter,navab}@cs.tum.edu
[2] Siemens Corporate Research, Princeton, NJ, USA
wolfgang.wein@siemens.com

Abstract. Recent US systems allow the real-time acquisition of volume data, either by freehand 3D techniques or novel transducer hardware. However, the acquisition of large volumes is limited by the field of view of the US transducer and anatomical scanning windows into the patient. Mosaicing of several 3D US scans has been proposed to generate large US volumes. Still, US imaging specific characteristics and artifacts make it challenging to create high quality mosaics. For many clinical cases, especially interventions, additional high quality CT data is available. In this paper we present a novel multi-variate, multi-modal 3D US registration and mosaicing approach, which reduces the effects of ultrasound imaging artifacts on mosaic quality, by incorporating information from co-registered CT.

1 Introduction

3D US imaging is becoming more and more frequently used in today's clinical workflows. Recent US systems allow direct real-time acquisition of 3D volume data. Either freehand 3D US, where a position and orientation (pose) sensor is attached to a conventional 2D US probe, or direct 3D acquisition capable US transducers are used. However the acquisition of large volume data is today still limited by the US transducer field of view and restricted scanning windows into the patient. Analog to 2D US image mosaicing, US volume mosaicing has been proposed to generate large US volumes from several acquired 3D US datasets.

Such 3D mosaic volumes of entire organs help to improve diagnostic confidence, e.g. by reducing scanning orientation induced artifacts, and transcend the role of ultrasound in radiology by providing self-contained 3D data for the reading room. For guidance of minimally-invasive interventions, such mosaics provide global anatomical reference, and allow for automatic mono- and multi-modal fusion with pre-operative data. US imaging specific artifacts, e.g. noise, occlusions, speckle, make it challenging to design robust similarity measures. By using the enriched spatial and artifacts reduced information of US mosaics the registration problem is becoming less ill-posed. Besides, comparison of 3D US at different stages before, during and after the procedure, allow to better assess treatment outcome. Additionally, 3D US mosaics could also be used as databases for generating simulated US images for training and education of physicians [1].

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 763–770, 2009.

1.1 Problem Statement

3D US mosaicing is a two step process. First the unknown registration parameters $[\hat{\boldsymbol{p}}_0, ..., \hat{\boldsymbol{p}}_N]$ of all individual volumes have to be computed. The registration problem is solved by minimization of a cost function

$$[\hat{\boldsymbol{p}}_0, ..., \hat{\boldsymbol{p}}_N] = \arg \min_{[\boldsymbol{p}_0, ..., \boldsymbol{p}_N]} \frac{\sum_{k=0}^{N} \sum_{l=0}^{N} \omega_{(k,l)} \cdot \sum_{\boldsymbol{v} \in \Omega} \Gamma \left(I_k[T^{-1}(\boldsymbol{p}_k) \cdot \boldsymbol{v}], I_l[T^{-1}(\boldsymbol{p}_l) \cdot \boldsymbol{v}] \right)}{\sum_{k=0}^{N-1} \sum_{l=k+1}^{N} \omega_{(k,l)}} \tag{1}$$

defined by the sum of the pair-wise cost functions Γ for all pairs (I_k, I_l) over all elements v in the overlap domain Ω, where $T^{-1}(\boldsymbol{p}_i)$ maps an element in overlap coordinates into the i-th volume. $\omega_{(k,l)}$ is a weighting factor for the contribution of the costs for pair (I_k, I_l) to the global costs. If one chooses $\omega_{(k,l)} = 1$ for all pairs the total cost is equal to the average of the individual costs. Another option is to choose $\omega_{(k,l)}$ equal to the pair-wise volume overlap, as proposed by Wachinger et al [2] to introduce a weighting favoring larger overlaps over smaller ones.

For solving the registration problem several approaches exist. They can be classified by the use of either direct or indirect cost functions, and pair-wise and simultaneous optimization approaches. Indirect approaches first pre-process the 3D US data to extract features for registration, in contrast direct approaches directly use the complete image intensity information. Pair-wise approaches decompose the global registration problem into several sequentially executed pair-wise registration problems, which are then used to compute the global registration by propagation of transformations. Simultaneous approaches directly minimize the cost function by simultaneously solving for all parameters and are more stable towards pair-wise local minima. However, they also greatly increase the computational effort per cost function evaluation. For symmetric cost functions, Eq.1 can be simplified to

$$\frac{\sum_{k=0}^{N-1} \sum_{l=k+1}^{N} \omega_{(k,l)} \cdot \sum_{\boldsymbol{v} \in \Omega} \Gamma \left(I_k[T^{-1}(\boldsymbol{p}_k) \cdot \boldsymbol{v}], I_l[T^{-1}(\boldsymbol{p}_l) \cdot \boldsymbol{v}] \right)}{\sum_{k=0}^{N-1} \sum_{l=k+1}^{N} \omega_{(k,l)}} \tag{2}$$

which reduces the evaluation costs significantly.

Once the registration problem is solved, the actual per pixel mosaic values have to be computed from the individual US intensities of different volumes. Both, the registration and mosaic steps are challenging to solve for US data due to the characteristics of US imaging itself. The intensity of the same anatomical region varies in different US scans for multiple reasons. Variations in direction of the US sound waves, give rise to occlusion and shadowing artifacts. Hypoechogenic regions can cause negative shadows in scan direction. Motion, either induced by pressure of the transducer onto the patient, internal organ motion, and of course breathing motion cause geometric variations in the images. Furthermore, US system internal signal processing, noise, speckle and operator dependent imaging control settings cause variations between different scans.

Several recent works have dealt with the problem of 3D US mosaicing. Gee et al [3] used a direct pair-wise registration approach. They reduce the complexity of the 3D/3D registration problem to a 2D/2D registration at volume intersection planes. Poon et al [4] proposed two block-based methods for improving the local alignment in the overlap volumes to compensate for the errors introduced by position trackers. Wachinger et al. [2] investigated registration strategies, pair-wise vs. simultaneous, for US mosaicing on a single phantom dataset. Simultaneous approaches proved to be superior to pair-wise ones, however resulting in high computational costs. Ni et al [5] chose an indirect pair-wise registration approach, using SIFT feature extraction. They evaluated their approach on an abdominal phantom and 3D US scans of human liver.

In this work we introduce a new simultaneous, direct, and multi-modal 3D US mosaicing approach. We chose not to use an indirect approach as reliable extraction of features from US data is inherently ill-posed in general for scans with varying field of view, scan direction, and unknown volume overlap. For many clinical scenarios a high quality CT scan of the patient is available. Our method uses the additional information from CT to improve the US/US registration by simultaneous registration of US to US and US to CT data. US/US registration problems caused by small overlap, lack of registration relevant features in the overlap volume, can be compensated by the additional information from CT. We use a variation of the method for CT/US fusion introduced by Wein et al [6] to estimate the registration parameters by simultaneously computing the registration of all US volumes to each other and to the CT scan. A simultaneous optimization was chosen instead of a pair-wise one to avoid accumulation of errors and improve overall stability of the registration. To reduce the computational costs we implemented the complete registration algorithm on GPU stream processors, greatly accelerating the multi-variate cost function evaluation compared to a CPU-based implementation.

2 Methods

Registration Framework. We use a GPU accelerated registration framework, supporting pair-wise, and simultaneous intensity based registration. Several, mono and multi-modal similarity measures (c.f. Table 1) have been implemented using OpenGL/GLSL[1]. We use global affine transformations, as rigid transformations are not sufficient to cover scale variations and shearing due to motion, scan conversion, compounding, shape of US field of view in the individual scans. A Downhill Simplex optimizer was chosen to optimize the similarity measure and determine the registration parameters. We added a transformation penalty term to the similarity measure to penalize physical impossible or unlikely transformations for US scans with a known pose guess from tracking. For US data with unknown pose, either a rough manual pre-positioning and relaxed penalty condition is used, or no transformation restriction at all.

[1] OpenGL - Open Graphics library http://www.opengl.org
GLSL - OpenGL's built-in high level shading language.

Table 1. Used similarity measures. Left colum, pair-wise, right column simultaneous version. (Row 1) n = 1 SAD - Sum of absolute differences, n = 2, SSD - Sum of squared differences (Row 2) NCC - Normalized Cross Correlation, (Row 3) LNCC - Local NCC in a k^3 neighbourhod.

$\frac{1}{N}\sum_{k=1}^{N} \lvert I_k - I_l\rvert^n$	$\frac{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}\sum_{\boldsymbol{v}\in\Omega}\lvert I_k - I_l\rvert^n}{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}\omega_{(k,l)}}$
$NCC(I,J) = \frac{\sum_{k=1}^{N}(I_k-\overline{I})(J_k-\overline{J})}{N\sigma_I\sigma_J}$	$\frac{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}\omega_{(k,l)}\sum_{\boldsymbol{v}\in\Omega} NCC(I_k,I_l)}{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}\omega_{(k,l)}}$
$LNCC(I,J,k) = \frac{1}{N}\sum_{nb=k}^{N} NCC(I,J)^k$	$\frac{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}\sum_{\boldsymbol{v}\in\Omega} LNCC(I_k,I_l)}{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}1}$

2.1 Multi-modal Mosaicing

The above described registration framework directly allows for mono-modal mosaicing of 3D US data. We incorporate the CT information into the registration algorithm by a variation of the method of Wein et al [6]. In every iteration the major US reflections are simulated from CT [7] and reconstructed into a cartesian volume [8] on the GPU. As similarity measure φ we chose the weighted sum of the multi-variate normalized cross correlation of all 3D US data, and real 3D US and simulated reflections (cf. Eq. 3).

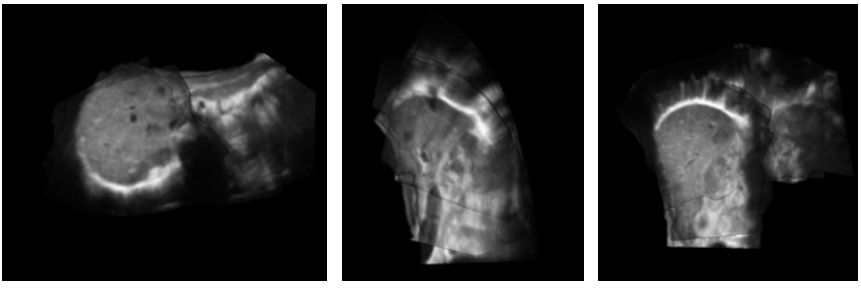
$$\varphi = \frac{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}\omega_{(k,l)}\cdot\sum_{\boldsymbol{v}\in\Omega}\mathbf{NCC}\left(US_k, US_l\right)}{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}\omega_{(k,l)}} + \alpha\frac{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}\omega_{(k,l)}\cdot\sum_{\boldsymbol{v}\in\Omega}\mathbf{NCC}\left(\theta(CT\ , US_k)), US_l\right)}{\sum_{k=0}^{N-1}\sum_{l=k+1}^{N}\omega_{(k,l)}} \tag{3}$$

where $\theta(CT\ , US_k)$ is the US simulation operator, simulating the k-th US volume from CT. For final refinements after registration we exchange the global NCC in φ with a Local NCC. Local NCC is not selected right from the start as NCC is stable enough for aligning the major features in US/US and CT/US data from the initial pose.

We chose this measure as the LC2 measure proposed by Wein et al [6] for CT/US fusion proved numerically instable on recent GPUs using GLSL due to lack of double precission floating point support. However, the simulated major reflections provide sufficient information for stable CT/US registration for US scans with large tissue interfaces.

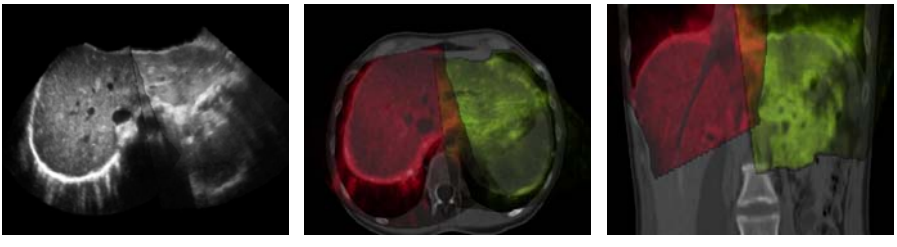
3 Experiments and Results

We evaluated our 3D US mosaicing framework for phantom and real clinical 3D freehand US datasets of human liver from multiple patients. Both mono- and multi-modal mosaicing were tested. The stability of the simultaneous US/US registration was evaluated on four 3D US scans of a baby phantom. The US probe was moved in cranio-caudal direction over the phantom, without a major view



(a) (b) (c)

Fig. 1. Axial along mosaic volume generated 2D MPRs from five freehand 3D US scans of human liver



(a) (b) (c)

Fig. 2. (a) Mosaic from two volumes with very small overlap in the middle. Mosaicing works only if CT information (b,c) is used to constrain the registration. Use of mono-modal approach will lead to breakup of the volume configuration.

direction change. Therefore the intensity variation in the scans is mostly due to noise. An optimization of translation and rotation parameters was performed 100 times with random offsets of +/- 10mm and +/- 5 degrees from a defined ground truth pose. SSD was used as similarity measure. Table 2a depicts the standard deviations of the translation and rotational parameters for the phantom and patient dataset.

The same stability study was performed for multiple real clinical datasets. Fig. 1 shows an example of an US mosaic created from four 3D freehand US scans of the liver region. For the clinical data NCC was selected as similarity measure. Table 3a depicts the results of the stability study. The results show that direct simultaneous registration performs as well on phantom as on real patient data. Tables 2b and 2b depict the performance of the framework for both mono and multi-modal registration. Table 2b shows an impressive speedup of more than 400 times of the GPU version versus a single-threaded CPU version for the same data and initial registration parameter settings. The used computer was equipped with an Intel Core Duo processor (2.6Ghz), 4Gbyte RAM and a NVIDIA Geforce GTX 280 with 1GByte of VRAM.

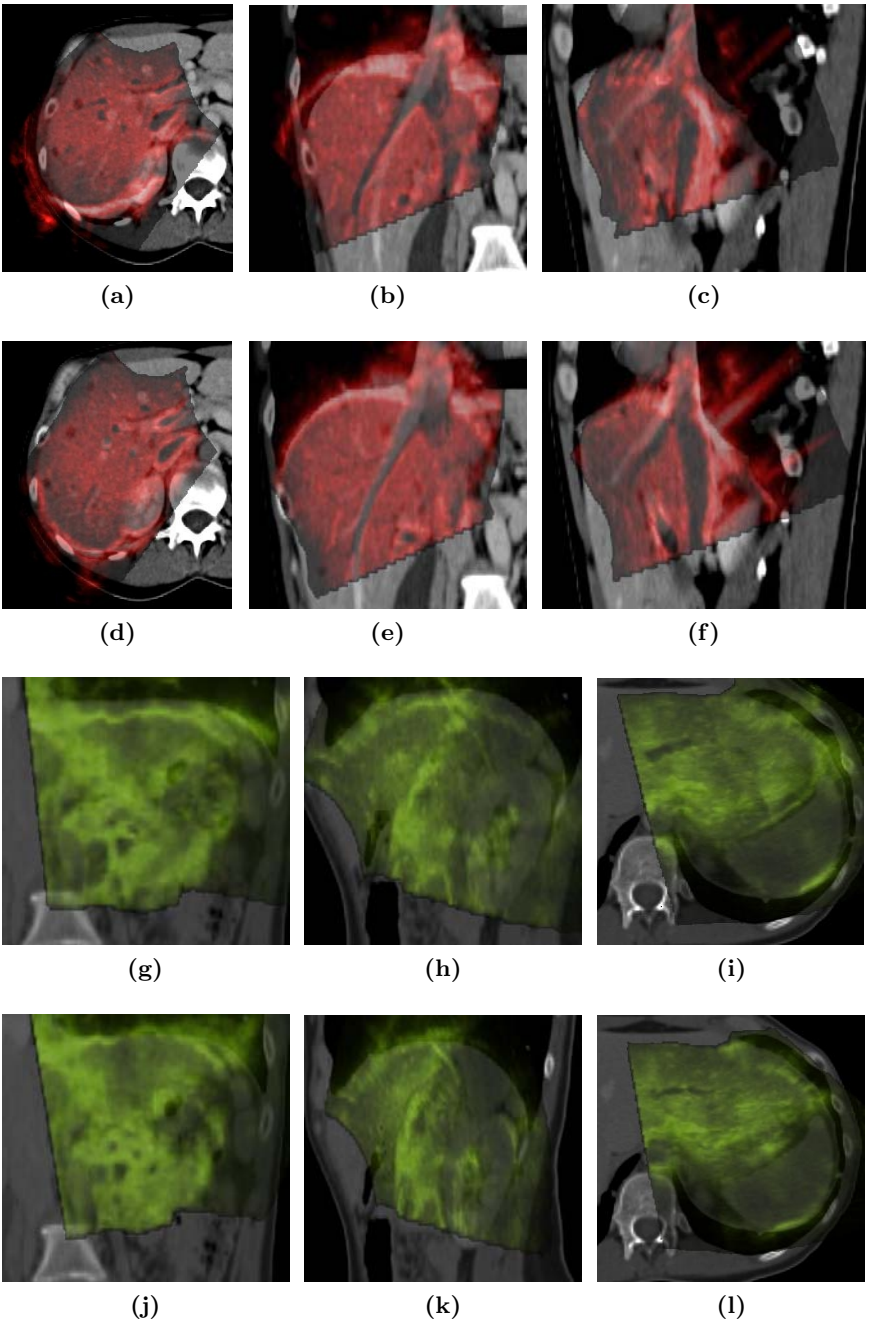


Fig. 3. Comparison of CT/US alignment before (a-c,g-i) and after registration (d-f,j-l) for the two US volumes with small overlap depicted in Fig. 2

Table 2. Simultaneous mono-modal registration experiment results for four US phantom scans. (a) Standard deviations of translation and rotation parameters for random study from defined ground truth pose. (b) GPU vs. CPU performance comparison for different volumes resolutions.

(a)

Vol.	tx [mm]	ty [mm]	tz [mm]	α [rad]	β [rad]	γ [rad]
1	3.2134	3.0474	3.1312	0.0183	0.0288	0.0204
2	3.1112	3.0478	3.2231	0.0198	0.0269	0.0212
3	3.1507	3.0117	3.1160	0.0198	0.0269	0.0206
4	3.1338	3.2669	3.3408	0.0176	0.0302	0.0227

(b)

	GPU [sec]	CPU [sec]
4 x 256^3	11.19231	> 36.000
4 x 128^3	9.71301	4006.34
4 x 64^3	7.10882	397.21

Table 3. Real patient data experient results. 256x256x106 voxels per volume (CT and US). (a) Standard deviations of translation and rotation parameters for random study. (b) Computational performance of simultaneous multi-modal registration for three and five volumes.

(a)

Vol.	tx [mm]	ty [mm]	tz [mm]	α [rad]	β [rad]	γ [rad]
1	3.2829	3.3165	3.5544	0.0212	0.0331	0.0230
2	3.2813	2.9114	2.9758	0.0173	0.0304	0.0232
3	3.2195	3.2878	3.8525	0.0176	0.0319	0.0259
4	3.2345	2.8773	3.1018	0.0170	0.0311	0.0243

(b)

Volume Config.	Time [sec]
1CT,2US	333.93
1CT,4US	1076.57

The multi-modal mosaicing was evaluated qualitatively by visual inspection of alignment of major anatomical regions, e.g. organs interfaces, vasculature inside the liver. The improvement of alignment before and after registration is clearly visible (see Fig. 3), but has to be evaluated detailed and qualitatively in further studies. We validated the improved registration for problematic configurations by a testcase with two US volumes with very small overlap and one CT scan (see Fig. 2). Using only US intensity information the registration fails and breaks the link of the US volumes. Using our proposed mosaicing method, the US volumes are registered to each other and to the CT data (see Fig. 3).

4 Discussion

In this work we have presented a novel, multi-modal and multi-variate, approach to 3D US mosaicing. The additional information about the patient anatomy from CT allows mosaicing of US scan configurations with very small pair-wise volume overlap and reduces problems caused by US intensity variations. We demonstrated that direct, simultaneous mono and multi-modal 3DUS mosaicing works on clinical data.In future work we will conduct an extensive evaluation of the proposed method on phantom and real clinical datasets. Physician defined landmarks in CT and US data will be used to assess the quality and correctness of the method. Our work also addressed the current major limitation of direct, simultaneous, multi-variate image registration by implementing the most expensive computations on stream processors of computer graphics hardware. Our

vision is to extend it to a general multi-modality, multi-variate image registration framework. The objective is to perform multi-modal registration not only for data from a pair of modalities but to incorporate all information into the registration framework.

Acknowledgments. This work was partially supported by Siemens Corporate Research and funding from the European Committee within the passport project - *http://www.passport-liver.eu/.*

References

1. Aiger, D., Cohen-Or, D.: Mosaicing ultrasonic volumes for visual simulation. IEEE Computer Graphics and Applications 20(2), 53–61 (2000)
2. Wachinger, C., Wein, W., Navab, N.: Registration strategies and similarity measures for three-dimensional ultrasound mosaicing. Academic Radiology 15, 1404–1415 (2008)
3. Gee, A., Treece, G., Prager, R., Cash, C., Berman, L.: Rapid registration for wide field of view freehand three-dimensional ultrasound. IEEE transactions on medical imaging 22(11), 1344–1357 (2003)
4. Poon, T., Rohling, R.: Three-dimensional extended field-of-view ultrasound. Ultrasound in Medicine & Biology 32(3), 357–369 (2006)
5. Ni, D., Qu, Y., Yang, X., Chui, Y.-P., Wong, T.-T., Ho, S.S.M., Heng, P.-A.: Volumetric Ultrasound Panorama Based on 3D SIFT. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 52–60. Springer, Heidelberg (2008)
6. Wein, W., Brunke, S., Khamene, A., Callstrom, M.R., Navab, N.: Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis 12, 577–585 (2008)
7. Kutter, O., Shams, R., Navab, N.: Visualization and GPU-accelerated simulation of medical ultrasound from CT images. Computer Methods and Programs in Biomedicine 94(3), 250–266 (2009)
8. Karamalis, A., Wein, W., Kutter, O., Navab, N.: Fast hybrid freehand ultrasound volume reconstruction. In: SPIE Medical Imaging, Orlando, Florida, USA (February 2009)

A Novel Method for Registration of US/MR of the Liver Based on the Analysis of US Dynamics

Sergiy Milko[1], Eivind Lyche Melvær[2], Eigil Samset[3], and Timor Kadir[4]

[1]IXICO, 2 Royal College Street, London, NW1 0NH, UK
[2]Centre of Mathematics for Applications, University of Oslo, 0316 Oslo, Norway
[3]Kongsberg SIM, 5 Bygdøy Allé , 0257 Oslo, Norway
[4]Siemens Molecular Imaging, 23-38, Hythe Bridge Street, Oxford, OX1 2EP, UK
{sergiym,eigilsa,eivindlm}@ifi.uio.no, timor.kadir@siemens.com

Abstract. Radiofrequency ablation of liver cancer is a minimally invasive alternative to open surgery. Typically, the preoperative planning is done on an MR (or CT) scan, while the intervention relies on ultrasound (US) guidance. Registration of intra-operative US and preoperative MR (or CT) would assist navigation and increase the confidence of RFA needle positioning. In this paper we present a novel method for registration of US and MR images of the liver. Hepatic vessels are extracted from 2D US by an algorithm that models US dynamics. It generates 2D probability maps representing hepatic vessels which are then combined into probability volumes. A multi-resolution registration framework performs registration of the pre-processed MR with two 3D vessel probability images. The accuracy, robustness and speed of the method were assessed by registering eight US/MR datasets. High robustness (86%) and reasonable accuracy (1.98°, 4.10mm), acceptable for the RFA clinical application, suggest that the method has a good potential for intra-operative use.

1 Introduction

Minimally invasive therapy is an increasingly important alternative to open interventions. Radiofrequency ablation (RFA) of liver cancer is a minimally invasive procedure that can be an alterative to surgery for inoperable tumors or can be combined with surgery in complex cases. An important step of the RFA treatment is insertion of an RF needle into the tumor. Typically, the insertion is performed under ultrasound (US) image guidance. However, US often produces images of poor quality for this task and can fail in depicting small lesions. MR and CT are complementary to US and can often image tumors missed by ultrasound as well as provide high resolution and high contrast image with a large field of view. By registering the preoperative images and the intra-operative US, it would be possible and clinically useful to include the visualization of such lesions to aid with accurate positioning of the needle.

The fusion of abdominal US and MR (or CT) images was a field of active research during the last decade. The problem is particularly complex due to fundamentally different physics involved in these imaging modalities. One strategy

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 771–778, 2009.

is to model the US from the MR (or CT) and register the real and simulated US images. In [1], Wein et al. developed a technique for simulating US from CT/CTA scans. The method is based on the empirical mapping between CT/CTA and US intensities, and modeling of US-wave propagation. A linear combination of the two was used to achieve the registration by optimizing an NCC-like similarity metric. The results are reported for the freehand 3D US and CTA datasets only. Roche et al. [2] utilized MR intensity and gradient information to simulate US images of the brain. Their Bivariate Correlation Ratio (BCR) similarity metric is based on consecutive optimization of a rigid transformation and simulated US.

To overcome the absence of functional dependency between MR (or CT) and US image intensities, another strategy is to map the datasets to an alternative, typically non-intensity, space - and register images in that space. In [3], Mellor and Brady generated local phase images from intensity US/MR and registered them using Mutual Information [4,5]. The method was only demonstrated on simulated brain US. Blackall et al. [6] produced probabilistic maps from US and MR images of the liver, where each pixel was assigned with a probability of belonging to the hepatic vessel tree. A registration of the two probabilistic representations was achieved by the Cross Correlation metric.

Recently, Milko et al. [7] reported a method for segmentation of liver US. The method differentiates between several anatomical structures, such as liver parenchyma and hepatic vessels. Importantly, the method analyzes not only the static US intensity, but also the dependency of pixel intensity on time. Authors demonstrated that modeling of time-dependent appearance of US has a clear benefit over pure intensity based segmentation approach. A preliminary investigation of method's feasibility for US/MR registration was introduced in [8]. However, only one pair of US/MR images was registered; no quantitative assessment of the registration was reported.

Motivated by the results in [7,8] and focused on the RFA application, we present a novel method for registration of preoperative MR to intra-operative liver US. The method is a significant advance over the ideas presented in [8]. It combines the analysis of US dynamics developed in [7] with a volume reconstruction algorithm and a multi-resolution registration framework. Namely, the information extracted from spatially tracked 2D US images is combined into a 3D volume that is registered with preoperative RFA planning. The accuracy, robustness and speed of the method were assessed on eight US/MR datasets.

2 Methods

2.1 Dynamic Texture Based Analysis of US

An excellent introduction to dynamic texture analysis for general computer vision is given in [9]. Its application to segmentation of liver US is reported in [7]. Here, we give a concise outline of the method from [7] focusing on the image registration problem. Let $\{I(t)|t = 1..\tau\}, I(t) \in R^m$ be a sequence of images. We call $\{I(t)\}$ a *(linear) dynamic texture* if it can be described by the system of two equations

$$\begin{cases} \bar{x}(t) = \sum_{i=1}^{k} A_i \bar{x}(t-i) + B\nu(t) \\ I(t) = \phi(\bar{x}(t)) + w(t) \ . \end{cases} \tag{1}$$

Here, $\bar{x}(t) \in R^n$ is a *state variable*. Its value at time point t linearly depends on previous k values, matrices $A_i \in R^{n \times n}, i = 1..k, B \in R^{n \times l}$ and a known independent and identically distributed (*IID*) sequence $\nu(t) \in R^l$. The equation for $I(t)$ involves a spatial filter $\phi : R^n \rightarrow R^m$ and another known *IID* sequence $w(t) \in R^m$. Following [7], we reduce (1) to a single equation describing the dynamics of individual US pixels. We consider the intensity function at each pixel p as a state variable $x_p(t) \in R^1$. Then $\{x_p(t)|t = 1..\tau\}$ is a time series - a sequence of 1D observations. By setting ϕ to identity transformation and all dimensions l, m, n to 1, the equation (1) can be reduced to

$$x_p(t) = A_1 x_p(t-1) + A_2 x_p(t-2) + \ ... \ + A_k x_p(t-k) + Bu(t) \ . \tag{2}$$

Equation (2) describes an *Auto Regressive (AR) process* characterized by linear coefficients $\{A_i, i = 1..k\}$ and B, and a residual noise $u(t)$. Respecting different assumptions regarding the type of noise for the US analysis, such as additive Gaussian [7] or multiplicative Rayleigh [10], we model $u(t)$ as a Gaussian with zero mean and unit variance due to its simplicity. Using least squares fitting, one can estimate the order k, coefficients $\{A_i, i = 1..k\}$ and B, that best explain the measurements $\{x_p(t)\}$. This was done by Schneider and Neumaier's `ARfit` algorithm [11].

In [7,8], a maximum a-posteriori (MAP) classifier was used to label each pixel with a tissue class, such as liver parenchyma or hepatic vessel, based on average (over time) pixel intensity and linear coefficients $\{A_i\}$ and B attributed to the pixel. As opposed to such "hard" classification, our method adopts a "soft" labeling where each pixel is assigned a probability of belonging to a certain tissue class. As described in [7], the probabilities were computed by a supervised MAP-classifier for liver parenchyma, hepatic and external (w.r.t. liver) vessels. Note, that hepatic- and external vessels can be separated by the level of "noisiness" (term B in (2)) and average pixel intensity. These two classes of vessels were identified as a basis for registration. To summarize: by analyzing a sequence of 2D US images acquired with a stationary held probe, we computed two probability maps - each reflecting the probability of US pixels to belong to one of the two vessel classes.

2.2 Reconstruction of Probability Volumes

The previous section outlined the analysis of a sequence of 2D images acquired with stationary US transducer. A number of such sequences were recorded, each producing two 2D probability maps for the two identified classes of vessels. Sequence acquisitions were performed at different positions of the probe covering a significant portion of the liver. The imaging planes were largely parallel, 2–3mm apart.

A median position and orientation of each US sequence was calculated based on the recordings of a magnetic tracking sensor attached to the probe. This spatial information was attributed to the 2D probability maps that were then weighted and combined into two probability volumes corresponding to two classes of vessels. Values of the 2D probability maps were weighted according to their geometrical position on the US image. A considerable area around the center of the US image was assigned the weight 1. The weights of pixels located further from the border of this area and towards the US edges gradually decrease to 0. This reflects higher confidence in structures observed in the center of the US image as opposed to boundaries (see Discussions). Probability values were multiplied with their weights and passed to the reconstruction step.

Each reconstruction was done by traversing a set of 2D probability maps twice. On the first traversal, a minimal bounding box containing the entire examined anatomy was determined. It was divided into a uniform grid with isotropic voxels having a resolution of 0.6mm. For all voxels, the probability and hit-counter values were initially assigned to zero. On the second traversal, probabilities were accumulated in a 3D volume. For each pixel p of the weighted 2D probability map a closest voxel v was determined based on the tracking and calibration data. The probability value of the voxel PR_v was updated and a hit-counter HC_v - incremented according to

$$PR_v := \frac{HC_v \times PR_v + PR_p}{HC_v + 1};\ HC_v := HC_v + 1$$

where PR_p is the probability value of the pixel p.

To avoid gaps in the reconstructed volume, we assume that 2D probability maps have a non-zero thinkness, which grows linearly as we move further from the head of the transducer. Thus, a 2D probability map is considered as a 3D "wedge" in the reconstructed volume and contributes to the intensity of all the voxels it intersects. The wedge thinkness was assigned to 1.75mm close to the transducer head and to 3.15mm at the opposite end. These values were derived empirically by visually inspecting the reconstructed volumes. Voxels having zero hit-counter ($HC_v = 0$) were excluded from the registration.

2.3 Registration

A proposed registration method operates with three 3D images: two reconstructed probability volumes corresponding to hepatic- and external vasculature, and a preprocessed MR scan. Note that probability volumes were reconstructed in the same coordinate space and are, therefore, registered. A preprocessing step for MR involves semi-automatic segmentation of major hepatic vessels as well as inferior vena cava and gallbladder. These two anatomical structures are typically visible on intra-operative US and have appearance similar to hepatic vessels. A volume produced on the MR preprocessing step is a binary mask containing ones in voxels belonging to the segmented structures and zeros elsewhere.

To define a formula for the registration metric, we introduce a general registration framework. Given a reference image I and a template image J, we look

for the rigid transformation T, which optimizes some similarity metric S. By x_k we denote the voxels of I. Then $i_k \equiv I(x_k)$ is an intensity value of I at x_k. Similarly, $j_k^{\downarrow} \equiv J(T(x_k))$, where the arrow $\downarrow$ marks the dependency on T. The position returned by $T(x_k)$ does not generally coincide with the voxel grid of J. Therefore, interpolation between the neighbour voxels of J is necessary. A partial volume (PV) interpolation was used as a tradeoff between speed and accuracy [2].

We chose a preprocessed MR image as a template (J) to respect the clinical context: registration of the pre-operative planning to the intra-operative US scan. Hence, a reference I is a vector image incorporating two probability volumes, and i_k consists of two components $i_k[hep]$ and $i_k[ext]$ representing probabilities of hepatic- and external vessel classes. A similarity metric can be defined as

$$S(T) = \sum_{x_k \in I^{\downarrow}} (a i_k[hep] + b i_k[ext]) j_k^{\downarrow} , \tag{3}$$

where $a, b \in [0..1]$ are linear coefficients reflecting the importance of two vessel classes for registration. Their values were chosen empirically as $a = 1$, $b = 0.5$ based on the observation explained in Discussions.

The similarity metric was maximized by the Powell optimization method. A multi-resolution registration framework was implemented. First, the images down-sampled to 30% of their original extent were registered. The obtained transformation was used to initialize the registration on the next step, where the registration of images down-sampled to 60% of the original dimension occurred. Finally, the original images were registered starting with the transformation computed on step two. We observed that registration of the original images improved the transformation only marginally, whereas its computational time was intense due to large dimensionality of datasets. Therefore, registrations for the first two resolutions only were used for numerical evaluation of the method.

The accuracy, robustness and speed of the method were assessed by registering eight pairs of 3D US/MR datasets. Each dataset consisted of three images: a preprocessed MR volume and two reconstructed probability volumes for liver US. "Ground truth" transformations were achieved manually by an expert registering the MR and reconstructed 3D US images. For this purpose, US volumes were reconstructed from the tracked 2D images by the algorithm described in the previous section.

We generated 200 perturbations from the ground truth by adding random rotations $\xi_r = (\xi_r^x, \xi_r^y, \xi_r^z)$ and translations $\xi_t = (\xi_t^x, \xi_t^y, \xi_t^z)$. The vectors ξ_r and ξ_t were chosen with random orientation, but constant magnitude: $\|\xi_r\| = 10°$ and $\|\xi_t\| = 10$mm. A total of $8 \times 200 = 1600$ automatic registrations were performed starting from perturbed initializations.

An automatic registration $R = \{\bar{\theta}, \bar{t}\} = \{\overline{(\theta_x, \theta_y, \theta_z)}, \overline{(t_x, t_y, t_z)}\}$ was considered *successful* if $\|\bar{\theta} - \bar{\theta}^*\| < \triangle\theta$ and $\|\bar{t} - \bar{t}^*\| < \triangle t$, where $R^* = \{\bar{\theta}^*, \bar{t}^*\}$ is a corresponding ground truth registration. The values for $\triangle\theta$ and $\triangle t$ were chosen as 3° and 6mm respectively to reflect the suitability for the RFA application.

The accuracy was measured on successful registrations achieved for each US/MR pair. Two errors were computed for every successful automatic registration $R = \{\bar{\theta}, \bar{t}\}$ as: $\delta_\theta = \|\bar{\theta} - \bar{\theta}^*\|$ and $\delta_t = \|\bar{t} - \bar{t}^*\|$. The mean $(\delta_\theta^{mean}, \delta_t^{mean})$ and median $(\delta_\theta^{median}, \delta_t^{median})$ rotation and translation errors were then calculated across all successful automatic registrations of each US/MR pair. The robustness was estimated as a percentage of successful registrations achieved by the method.

3 Results

Table 1 shows the robustness, accuracy, and speed of the registration method tested on eight pairs of US/MR images. Recall that 200 registrations were performed for each image pair. Therefore, a success rate of 0.5% refers to a single successful registration. Mean $(\delta_\theta^{mean}, \delta_t^{mean})$ and median $(\delta_\theta^{median}, \delta_t^{median})$ errors were computed on successful registrations only.

Table 1. Accuracy, robustness and speed for registering eight US/MR datasets. Note, that $(\delta_\theta^{mean}, \delta_t^{mean})$ and $(\delta_\theta^{median}, \delta_t^{median})$ errors were computed on successful registrations only. Robustness was measured as a success rate (%). The value 0.5% corresponds to one registration out of two hundred.

Dataset	$\delta_\theta^{mean}, ^\circ$	δ_t^{mean}, mm	$\delta_\theta^{median}, ^\circ$	δ_t^{median}, mm	Time,sec	Success rate,%
1	1.16	5.27	1.16	5.27	569	81.0
2	2.75	5.78	2.75	5.77	473	87.0
3	2.05	3.40	2.05	3.39	425	87.5
4	1.39	3.47	1.37	3.43	467	86.0
5	1.67	5.02	1.66	4.99	508	81.5
6	2.89	4.60	2.89	4.60	592	87.0
7	2.11	3.36	2.11	3.36	536	83.0
8	1.73	2.13	1.73	2.14	346	92.0
Avg. across all	**1.98**	**4.10**	**2.01**	**3.47**	**525**	**86.0**

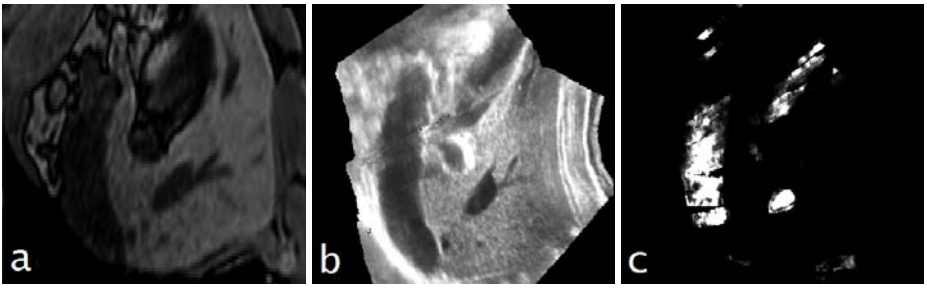


Fig. 1. Registration of the US/MR datasets: a - original MR; b - reconstructed 3D US; c - probability volume corresponding to hepatic vessels

An example of successfully registered US/MR images of the liver is presented on fig. 1, where cross sections through the original US/MR and a probability volume corresponding to hepatic vessels are shown.

4 Discussion and Conclusion

Table 1 demonstrates that the proposed registration method has a good potential for the RFA clinical application due to its high robustness (86%) and acceptable rotation (1.98°) and translation (4.10mm) errors, averaged across all 1600 registrations. The method is somewhat similar to the algorithm proposed in [6], where manually segmented MR was registered to a sparse set of 2D vessel probability maps computed on the US intensity. We believe that our approach has several advantages. Modeling the dynamics of liver US enables capturing higher number of hepatic vessels and external (w.r.t. liver) vascular structures than the analysis of static US intensity only [7]. This statement is particularly valid for vessels distant from the US probe that have average (over time) intensity similar to the adjacent liver parenchyma. The reconstruction of 3D probability images from a set of 2D probability maps is beneficial as it compensates for re-occurrences of the same anatomy on more than one 2D image. Indeed, since the US acquisition was performed with a hand held probe with relatively small distance between imaging planes (2–3mm), their intersection often occurred. If a sparse set of 2D probability maps was used, re-occurring vessels would influence the registration more than vessels encountered only once.

Multi-resolution registration framework is a popular choice when the similarity metric is not sensitive to down-sampling of the registered images. This holds for the proposed metric computed on US probability volumes and a preprocessed MR. Computational efficiency achieved by the multi-resolution approach makes it attractive for intra-operative use.

The major limitation of our method, as well as the method described in [6], is a necessity for manual preprocessing of the MR. However, it is mitigated by the RFA application - a clinical context of both studies. A preoperative planning for the RFA often includes delineation of critical structures, such as tumors and large adjacent vessels, on the MR. Availability of this information can make the registration method fully automatic. The algorithm does not compensate for non-rigid deformations and is, therefore, sensitive to breathing. The RFA treatment is typically conducted under general anesthesia. It is a common practice to suspend subject's breathing during the needle insertion phase. Therefore, rigid registration is adequate at this stage of the RFA procedure.

We believe that by utilizing probabilities rather than "hard" segmentation as described in [7,8], our method gives better alignment and is more robust to errors in US segmentation. The registration accuracy is, however, limited by the accuracy of the MR segmentation. To address this issue, registration of US probability volumes directly to MR intensities is considered a future work.

A selection of linear coefficients $a = 1$ and $b = 0.5$ in (3) respects the higher importance of hepatic vessels for registration than either external (w.r.t. liver)

vessels or the gallbladder. Due to their larger size, the gallbladder and inferior vena cava can force the registration to drift away from the correct solution unless down-weighted.

A confidence mask applied to the US probability maps reflects higher confidence in structures observed in the center of the US image, as opposed to boundaries. For example, a US region adjacent to the transducer contains compressed and, therefore, distorted structures; a boundary region distant from the probe receives attenuated US signal and is often noisy.

In this paper we proposed a novel method for registration of preoperative MR and intra-operative liver ultrasound based on modeling of US dynamics. Quantitative assessment demonstrated high robustness of the method and accuracy acceptable for the RFA application. Despite the necessity of manual preprocessing step for pre-operative MR, the method benefits from frequent availability of such information for the RFA treatment.

References

1. Wein, W., Röper, B., Navab, N.: Integrating Diagnostic B-mode Ultrasonography into CT-based Radiation Treatment Planning. IEEE Trans. on Med. Imag. 26(6), 866–879 (2007)
2. Roche, A., Pennec, X., Malandain, G., Ayache, N.: Rigid Registration of 3D Ultrasound with MR Images: a New Approach Combining Intensity and Gradient Information. IEEE Trans. on Med. Imag. 20(10), 1038–1049 (2001)
3. Mellor, M., Brady, M.: Fluid Registration of Ultrasound Using Multi-scale Phase Estimates. In: Proc. of BMVA 2005 (2005)
4. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality Image Registration by Maximization of Mutual Information. IEEE Trans. on Med. Imag. 16(2), 187–198 (1997)
5. Viola, P., Wells, W.M.: Alignment by Maximization of Mutual Information. Int. Journal of Comp. Vision. 24(2), 137–154 (1997)
6. Blackall, J.M., Penney, G.P., King, A.P., Hawkes, D.J.: Alignment of Sparse Freehand 3-D Ultrasound with Preoperative Images of the Liver Using Models of Respiratory Motion and Deformation. IEEE Trans. on Med. Imag. 24(11), 1405–1416 (2005)
7. Milko, S., Samset, E., Kadir, T.: Segmentation of the Liver in Ultrasound: a Dynamic Texture Approach. Int. Journal of Comp. Assist. Radiol. and Surgery 3(1-2), 143–150 (2008)
8. Milko, S., Samset, R., Kadir, T.: A Novel Dynamic Texture Based Approach for Segmentation and Registration of Liver Ultrasound. In: Proc. of IGSTI 2008 (2008)
9. Doretto, G., Chiuso, A., Wu, Y.N., Soatto, S.: Dynamic Textures. Int. Journal of Comp. Vision 51(2), 91–109 (2003)
10. Boukerroui, D.: Velocity Estimation in Ultrasound Images: A Block Matching Approach. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 586–598. Springer, Heidelberg (2003)
11. Schneider, T., Neumaier, A.: Algorithm. ARfit - A Matlab Package for the Estimation of Parameters and Eigenmodes of Multivariate Autoregressive Models. ACM Trans. on Math. Software 27, 58–65 (2001)

Alignment of Viewing-Angle Dependent Ultrasound Images*

Christian Wachinger and Nassir Navab

Computer Aided Medical Procedures (CAMP), TUM, Munich, Germany
{wachinge,navab}@cs.tum.edu

Abstract. We address the problem of the viewing-angle dependency of ultrasound images for registration. The reflected signal from large scale tissue boundaries is dependent on the incident angle of the beam. This applies an implicit weighting on the ultrasound image, dependent on the viewing-angle, which negatively affects the registration process, especially when utilizing curved linear transducers. We show that a simple reweighting of the images, considering a common physical model for ultrasound imaging, is not feasible. We therefore introduce a new matching function, separating reflectivity and scattering regions, which are the results of two different types of physical interactions of the ultrasound beam with the tissue. We use the local phase for identifying regions of reflectivity, and consider it as one part of our matching function, combining feature- and intensity-based aspects. First experiments provide good results for this novel registration approach.

1 Introduction

The fusion of several ultrasound images for the creation of a combined image with a larger field-of-view, referred to as ultrasound stitching, mosaicing or panorama imaging, is of clinical interest [1] and topic of on-going research. In the past, a variety of methods were proposed ranging from rigid to deformable [2], from intensity- to feature-based [3], and from pair- to groupwise [4]. While feature-based approaches may be an option for getting close to the correct alignment, the final alignment is best achieved with groupwise intensity-based methods [4].

The major drawback of intensity-based methods is the missing overlap invariance of the similarity measures, favoring a total overlap of the images [4]. The reasons for this phenomenon have not yet been investigated. They could either be rooted in the similarity measures themselves, like Cahill *et al.* [5] addressed, or in the ultrasound images. We show that the viewing-angle dependency of ultrasound images not only causes angle dependent artifacts like shadow, but also puts an implicit weighting on the images, favoring a total overlap.

To address this issue we first evaluate methods for reweighting ultrasound images, making them viewing-angle independent. However, currently used ultrasonic imaging models are not accurate enough, as we will show, for allowing

* This work was partly funded by the European Project "PASSPORT", ref. number 223904. The authors thank Andreas Keil and Reviewer 1 for valuable contributions.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 779–786, 2009.

such a reweighting. We instead propose a new ultrasound matching function, consisting of one term, measuring the alignment of reflectivity structures, and a second term, measuring the alignment of the remaining texture. This new matching function has advantages for US-US, as well as, US-CT registration.

2 Viewing Angle Dependency

Structures, visible in ultrasound images, are the result of two different types of interaction of the ultrasound beam with the tissue: reflection and scattering. Reflection occurs at large scale tissue boundaries and is viewing-angle dependent. Scattering is caused by microscopic tissue inhomogeneities and provides the internal texture of the organs. Scattering is to a certain extent dependent on the direction of insonification because the ultrasound PSFs are not spherically symmetric. However, since we are not able to match single scattering responses from the resolution cells, it is more appropriate to consider entire scattering regions, and the echogeneity of these regions does not change with the viewing-angle.

A common physical model for the reflection, as it is described in the textbooks of Hedrick *et al.* [6] or Zagzebski [7], and further used by Wein *et al.* [8], is to calculate the reflected signal $R(\mathbf{x})$ as

$$R(\mathbf{x}) = \rho(\mathbf{x}) \cdot I_i(\mathbf{x}) \cdot (\cos \phi(\mathbf{x}))^m \tag{1}$$

with the incident angle $\phi(\mathbf{x})$ at position $\mathbf{x}$, and the incident intensity I_i. The reflection coefficient calculates as

$$\rho(\mathbf{x}) = \left(\frac{z(\mathbf{x}) - z(\mathbf{x} - \Delta\mathbf{d})}{z(\mathbf{x}) + z(\mathbf{x} - \Delta\mathbf{d})} \right)^2 \tag{2}$$

with the acoustic impedance z and $\Delta\mathbf{d}$ the distance between scan line points. The exponent m models the heterogeneity of the tissue interface, ranging from rather specular reflections to diffuse reflectors. Typically, strong reflections can be observed at the diaphragm and the kidney surface.

To illustrate the problems, that affect the registration process when working with viewing-angle dependent ultrasound images, we create two artificial images, acquired from a curved linear probe, see Figure 1. The images show a region of high reflectivity, having a cosine intensity profile, and a scattering region. Rayleigh distributed noise is added to the images (log-compressed) to simulate the speckle noise of the ultrasound images [9]. The correct alignment of the two images, which can be acquired with lateral displacement, is to overlap the scattering regions. However, as we show in the similarity plot of Figure 1, similarity measures like CD_2 [9] favor a total overlap of the images. The reason is the dominance of the reflectivity structure, leaving only a local minimum at the desired position.

2.1 Reweighting Ultrasound Images

A first attempt, when thinking about the problem, is to identify structures of high reflectivity, calculate the incidence angle of the beam, and use the ultrasonic

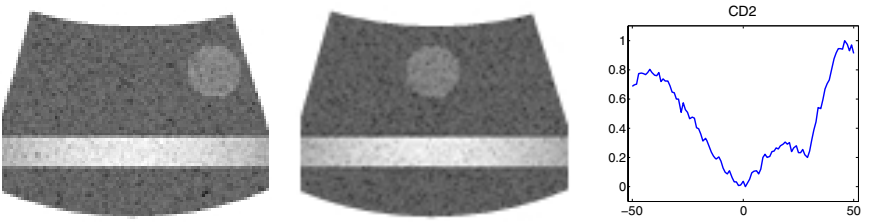


Fig. 1. Artificial ultrasound images showing reflection, scatter, and speckle. Similarity plot using CD_2 [9] along lateral direction

imaging model as shown in Equation (1) to reweight the image. The result would be the creation of normalized ultrasound images being less or not at all viewing angle dependent. In the case of our example, the reflectivity region would result in a constant similarity value for translations in lateral direction, so that the scattering region would dominate the registration process.

In the following, we will investigate if the presented ultrasound model is accurate enough for reweighting the images. We acquired ultrasound B-mode and RF images with a curved linear transducer in a water bath with just one object in it, having a diffuse surface, usually used for single wall calibration, see Figure 2. The reflection coefficient $\rho(\mathbf{x})$ is the same for every beam, as we have the common water-object boundary. The incident intensity I_i is dependent on the attenuation of the beam along its way to the boundary. The attenuation coefficient, which includes absorption and scattering effects, is $2.2 \cdot 10^{-3} \frac{\text{dB}}{\text{cm} \cdot \text{MHz}}$ for water. In comparison, the attenuation coefficients for Fat ($0.6 \frac{\text{dB}}{\text{cm} \cdot \text{MHz}}$), Liver ($0.9 \frac{\text{dB}}{\text{cm} \cdot \text{MHz}}$), and Kidney ($1.0 \frac{\text{dB}}{\text{cm} \cdot \text{MHz}}$), are about three orders of magnitude larger. Further, the length of the beam in the middle and the one on the side vary about 2 cm (1.93 exactly). This leads to an attenuation of $22.8 \cdot 10^{-3}$ dB by using an acquisition frequency of 2.6 MHz and considering also the way back to the transducer. Since the ratio of change of intensity $\frac{I_1}{I_0}$ is close to 1 for values close to 0 dB, we can neglect the effect of attenuation for our experiment.

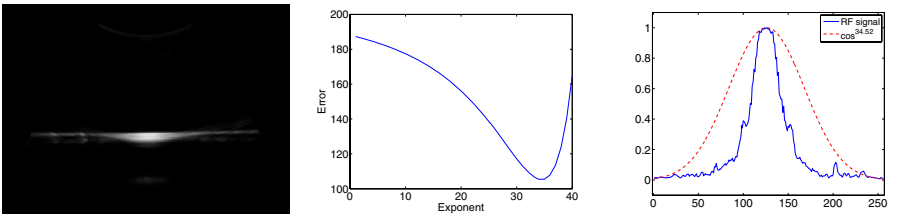
The received signal, therefore only depends on the incident angle of the beam. The subsequent cost function has to be minimized in order to find the optimal exponent m, characterizing best the type of the interface

$$\mathcal{C}(m) = \int \left(\frac{|R(\mathbf{x})|}{\cos^m(\phi(\mathbf{x}))} - \rho(\mathbf{x}) \cdot I_i(\mathbf{x}) \right)^2 d\mathbf{x} \tag{3}$$

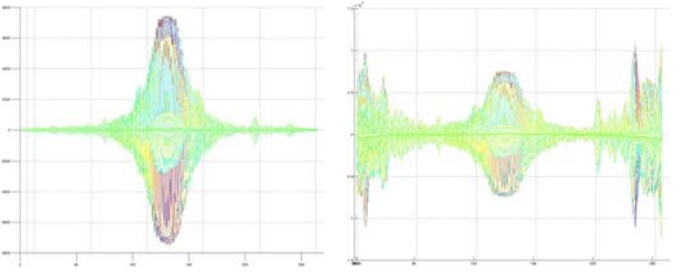
$$\frac{d\mathcal{C}(m)}{dm} = \int 2 \left(\frac{|R(\mathbf{x})|}{\cos^m(\phi(\mathbf{x}))} - \rho(\mathbf{x}) I_i(\mathbf{x}) \right) |R(\mathbf{x})| \cdot \ln(\cos(\phi(\mathbf{x}))) \cdot \cos^{-m}(\phi(\mathbf{x})) d\mathbf{x}$$

with $\rho(\mathbf{x}) \cdot I_i(\mathbf{x})$ being constant for all $\mathbf{x}$ on the surface and corresponding to the measured reflectivity of the middle beam, and $|R(\mathbf{x})|$ the absolute value of the received signal.

We plot the cost function in Figure 2(b), where a gradient descent optimization with the above stated gradient of the cost function leads to $m = 34.52$. In Figure 2(c), we show the intensity profile of the RF data, together with



(a) B-Mode US image (b) Cost Function (c) Intensity Profile and Best Approximation



(d) RF Profile (x-axis: lateral) (e) Reweighted RF Profile

Fig. 2. Reweighting experiment on ultrasound RF data

the optimal cosine variant. The advantage of working with RF data is that no further processing steps or filters are applied on the data, making a meaningful evaluation of the model possible. When working with B-mode images, especially the log-compression has to be considered, turning the multiplicative cosine term of Equation (1) into an additive one. In Figure 2(e), we show the reweighted RF signal with $\cos^{34.52}$. As can be seen, the result is not satisfying because instead of having a constant intensity profile, we have a "W" shaped one. For other interfaces it may be possible to find an exponent m so that reweighting works, however, this would only characterize this specific type of interface and not be generally applicable. We conclude that the presented ultrasonic imaging model in Equation (1) is usually not accurate enough to allow for reweighting ultrasound images, in order to make them viewing-angle independent.

This finding also affects a recently introduced algorithm for US-CT registration [8], which bases upon the simulation of ultrasound images, applying this ultrasonic imaging model. The the next section, we introduce a new ultrasound matching scheme, usable for US-US and US-CT registration, to address this issue.

3 A New Ultrasound Matching Scheme

In this section we present a novel matching scheme for ultrasound images, dealing with the viewing angle dependency, discussed in the last section. As we

concluded, a correction of the angle dependency of the images is with the current model not feasible. Our approach is therefore to separate reflection and scatter, which are the results of two different physical interactions of the beam with the tissue, combined in one image. In the following, we first present the method that we use for the crucial identification of regions of reflectivity in the ultrasound images, and then detail the new matching function.

3.1 Reflectivity Regions

We use local phase information for identifying regions of reflectivity because it provides us with structural information independent of the brightness and contrast [10]. This independence is very important for the extraction of reflectivity regions, since the absolute intensity of the reflection varies with the incident angle. The local phase was already used by Mulet-Prada and Noble [11] for boundary detection in echocardiography. Recently, Hacihaliloglu *et al.* [12] achieved good results in segmenting bone in ultrasound images employing the local phase. Mellor and Brady [13] apply mutual information on local phase images for multi-modal image registration. Grau *et al.* [10] use the local phase for aligning 3D echocardiographic sequences.

For 1-D signals the phase is constructed from the original signal and its Hilbert transform. Different approaches exist to extend this concept to N-D. In the following, we use the *monogenic signal* introduced by Felsberg *et al.* [14]. It uses a generalization of the Hilbert transform, the Riesz transform, to calculate phase information in N-D. The image is filtered by N filters, which are given in the Fourier domain by

$$H_i(f_1, \ldots, f_N) = \frac{f_i}{\sqrt{\sum_{j=1}^{N} f_j^2}} \tag{4}$$

with $f_1, \ldots, f_N$ the Fourier domain coordinates. We follow [10] in applying log-Gabor filters prior to the calculation of the monogenic signal of the image to extract frequency and spatial localization. The monogenic signal provides us with information about the local phase $P(x)$ each pixel x. We threshold the phase image with a value τ to obtain a binary mask, indicating reflectivity regions. Throughout all the experiments we use $\tau = 0.7$, showing that this is not a crucial parameter. Alternatively, the local phase values could directly be used as weights, resulting in a fuzzy like mask.

3.2 Matching Function

Since the local phase image provides us with structural information independent of brightness, building a matching function upon this like in [10] would already significantly reduce the effects of the viewing-angle dependency on the registration. However, this also leads to the following drawbacks: First, we would be completely dependent on the outcome of the algorithm calculating the local phase. And second, texture information in the form of scattering regions is not integrated. Focusing on boundaries seems totally fine for echcardiographic

applications, where not much texturing exists, but for other applications, like abdominal ones, this is not optimal. Consequently, the matching function we propose considers both, feature- and intensity-based aspects, by using reflection and texture.

Considering two images u and v, an image grid Ω, regions of reflectivity $\Gamma_u = \{x \in \Omega | P_u(x) > \tau\}$ and $\Gamma_v = \{x \in \Omega | P_v(x) > \tau\}$, further $\Gamma = \Gamma_u \cup \Gamma_v$, the matching function is

$$\mathcal{M}(u,v) = \mathcal{M}^T_{\Omega \setminus \Gamma}(u,v) + \lambda \cdot \mathcal{M}^R_{\Omega}(P_u, P_v) \tag{5}$$

with $\mathcal{M}^T_{\Omega \setminus \Gamma}$ measuring the similarity of the texture regions, excluding reflectivity regions $\Omega \setminus \Gamma$, and $\mathcal{M}^R_{\Omega}(P_u, P_v)$ measuring the similarity of the phase images on the whole grid Ω. The matching function can be easily extended to the multivariate case by using the multivariate similarity framework proposed in [4], summing up pairwise terms.

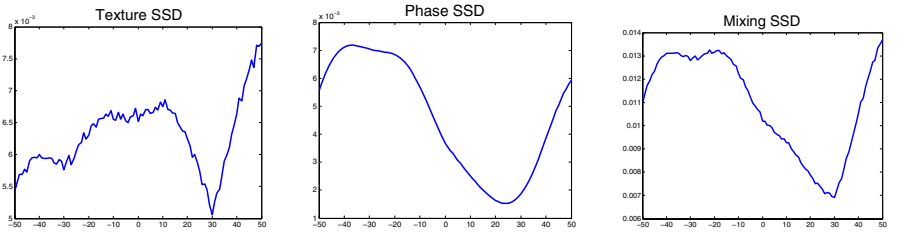
Suitable similarity measures for $\mathcal{M}^T$ are, next to standard measures like sum of squared differences (SSD), normalized cross-correlation (NCC), and mutual information (MI) assuming a Gaussian distributed noise, the similarity measures proposed by Cohen and Dinstein [9], assuming a Rayleigh distributed noise. Especially CD_2 was reported to have good performance [15]. Suitable similarity measures for $\mathcal{M}^R$ are once again the standard measures, and the similarity measure proposed in [10], taking the cosine of the phase difference.

The disadvantage, when choosing two different similarity measures, is that the correct selection of the weighting term λ is important, since it has to map the similarity score into a comparable range. Even when working with normalized images, the outcomes of the similarity measures are not directly comparable, so that we perform the experiments with the same similarity measures for both terms and simply set $\lambda = 1$.

This matching framework is also directly applicable to the US-CT registration [8], where texture and reflection parts are simulated separately and finally combined during the registration. In our matching approach, reflectivity and texture regions are extracted from the ultrasound image, and can directly be matched with the simulated ones.

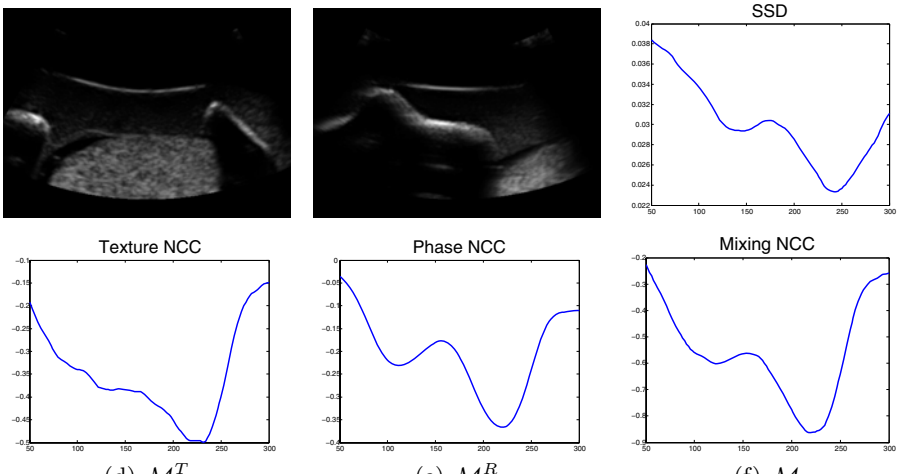
4 Experiments

In Figure 3, we show the results of the new matching function, using SSD as similarity metric for texture and reflection, for the data set illustrated in Figure 1. Comparing the similarity plots for CD_2 and $\mathcal{M}^T$, which both use the texture, we can see the improvement of not considering the reflectivity region Γ in the calculation. The plot of the term $\mathcal{M}^R$, measuring the similarity of the phase images is very smooth, however, does not indicate the correct alignment at 30. Combining both curves, $\mathcal{M}$, leads to an accurate cost function with wider capture range than $\mathcal{M}^T$.



(a) $\mathcal{M}^T$ (b) $\mathcal{M}^R$ (c) $\mathcal{M}$

Fig. 3. Similarity plots for images from Figure 1 along lateral direction using SSD



(d) $\mathcal{M}^T$ (e) $\mathcal{M}^R$ (f) $\mathcal{M}$

Fig. 4. Ultrasound images acquired from laterally displace positions. Similarity plots for SSD and the different parts from our new matching scheme.

We also conducted experiments for US-US registration on ultrasound images acquired with a curved linear transducer, see Figure 4. The images are acquired from an ultrasound phantom, with a true displacement of 88.0 mm. The bony structure on the left side of the first image is depicted in the middle of the second image. The similarity plots for SSD, NCC, and CD_2 are comparable, with SSD indicating the correct alignment at 96.8 mm. The combined cost function indicates the correct alignment at 88.8 mm, with NCC as similarity measure for both terms.

We performed further experiments on a data set, similar to the one presented by Penney *et al.* [16], simulating the US-CT registration problem. An illustration of the images together with the good results of our matching approach can be found in the supplementary material [1].

[1] http://www.webcitation.org/5hEZWaajc

5 Conclusion

The alignment of ultrasound images is complicated by the viewing-angle dependent nature of the images. After an analysis of the effects on the registration process, we presented a method for reweighting ultrasound images. However, the commonly applied imaging model for ultrasound was shown to be not accurate enough, for allowing a correction of the viewing-angle dependency. Instead, we introduced a new matching scheme for the alignment of ultrasound images, which can be used for US-US as well as US-CT registration. It incorporates the usage of feature- and intensity-based metrics into one global measure. We evaluated the new registration approach on various data sets, with good results.

References

1. Henrich, W., Schmider, A., Kjos, S., Tutschek, B., Dudenhausen, J.W.: Advantages of and applications for extended field-of-view ultrasound in obstetrics. Archives of Gynecology and Obstetrics 268, 121–127 (2003)
2. Poon, T., Rohling, R.: Three-dimensional extended field-of-view ultrasound. Ultrasound in Medicine and Biology 32(3), 357–369 (2005)
3. Ni, D., Qu, Y., Yang, X., Chui, Y.-P., Wong, T.-T., Ho, S.S.M., Heng, P.-A.: Volumetric ultrasound panorama based on 3D SIFT. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 52–60. Springer, Heidelberg (2008)
4. Wachinger, C., Wein, W., Navab, N.: Three-dimensional ultrasound mosaicing. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 327–335. Springer, Heidelberg (2007)
5. Cahill, N., Schnabel, J., Noble, J., Hawkes, D.: Revisiting overlap invariance in medical image alignment. In: MMBIA (2008)
6. Hedrick, W.R., Hykes, D.L., Starchman, D.E.: Ultrasound Physics and Instrumentation. Mosby, 4th edn. (2004)
7. Zagzebski, J.: Essentials Of Ultrasound Physics, 1st edn., Mosby (1996)
8. Wein, W., Brunke, S., Khamene, A., Callstrom, M.R., Navab, N.: Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis 12, 577–585 (2008)
9. Cohen, B., Dinstein, I.: New maximum likelihood motion estimation schemes for noisy ultrasound images. Pattern Recognition 35(2), 455–463 (2002)
10. Grau, V., Becher, H., Noble, J.: Registration of multiview real-time 3-d echocardiographic sequences. IEEE Transactions on Medical Imaging 26(9) (September 2007)
11. Mulet-Parada, M., Noble, J.: 2D+ T acoustic boundary detection in echocardiography. Medical Image Analysis 4(1), 21–30 (2000)
12. Hacihaliloglu, I., Abugharbieh, R., Hodgson, A.J., Rohling, R.: Bone segmentation and fracture detection in ultrasound using 3D local phase features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 287–295. Springer, Heidelberg (2008)
13. Mellor, M., Brady, M.: Phase mutual information as a similarity measure for registration. Medical Image Analysis 9(4), 330–343 (2005)
14. Felsberg, M., Sommer, G.: The monogenic signal. IEEE Transactions on Signal Processing 49(12), 3136–3144 (2001)
15. Boukerroui, D., Noble, J.A., Brady, M.: Velocity estimation in ultrasound images: a block matching approach.. Inf. Process Med. Imaging 18, 586–598 (2003)
16. Penney, G., Griffin, L., King, A., Hawkes, D.: A novel framework for multi-modal intensity-based similarity measures based on internal similarity. SPIE 6914 (2008)

MR to Ultrasound Image Registration for Guiding Prostate Biopsy and Interventions

Yipeng Hu[1], Hashim Uddin Ahmed[2], Clare Allen[3], Doug Pendsé[3], Mahua Sahu[2], Mark Emberton[2], David Hawkes[1], and Dean Barratt[1]

[1] Centre for Medical Image Computing, University College London, London, UK
[2] Department of Urology, Division of Surgery & Interventional Science, University College London, London, UK
[3] Department of Radiology, University College London Hospital (UCLH), London, UK

Abstract. A method is described for registering preoperative magnetic resonance (MR) to intraoperative transrectal ultrasound (TRUS) images of the prostate gland. A statistical motion model (SMM) of the prostate is first built using training data provided by biomechanical simulations of the motion of a patient-specific finite element model, derived from a preoperative MR image. The SMM is then registered to a 3D TRUS image by maximising the likelihood of the shape of an SMM instance given a voxel-intensity-based feature, which represents an estimate of normal vector at the surface of the prostate gland. Using data acquired from 7 patients, the accuracy of registering T2 MR to 3D TRUS images was evaluated using anatomical landmarks inside the gland. The results show that the proposed registration method has a root-mean-square target registration error of 2.66 mm.

1 Introduction

Prostate cancer is a major international health problem, particularly affecting men in the Western World. Traditional treatment strategies for localised prostate cancer involve either radical treatment of the whole gland - for example, by surgical excision or radiotherapy - or pursuing an active surveillance/watchful waiting programme in which intervention is delayed in favour of monitoring the patient for signs of disease progression. Motivated by the potential for a reduced risk of side-effects compared with conventional treatments, there has recently been growing interest in techniques which enable the targeted treatment of prostate cancer and minimise damage to vulnerable structures, such as nerves [1]. However, accurate mapping of cancer foci and accurate therapy delivery are of critical importance in such approaches.

Transrectal ultrasound (TRUS) imaging remains the most accessible and practical means for guiding needle biopsy and therapeutic interventions. However, has poor accuracy for visualisation of cancer foci within the prostate means conventional (B-mode) TRUS imaging provides very limited information on the spatial location of tumours. Consequently, TRUS images are not suitable for disease mapping, and performing accurate, targeted biopsy. Instead, current standard of care is to use TRUS to simply inform the operator that the needle is in the prostate and which side it is in. Recent advances in functional and structural MR imaging techniques for localising

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 787–794, 2009.

and characterising prostate cancer have led to sensitivities and specificities that are now sufficiently high to be clinically useful for guiding biopsy needles and for targeting localised therapy in a more precise manner [2]. However, the ability to fuse anatomical and pathological information on tumour location, derived from MR images or a previous biopsy, with TRUS images obtained during a procedure remains a significant technical challenge.

In this paper, a new method is presented for non-rigid registration of MR images and 3D TRUS images, which compensates for gland motion and is sufficiently fast for intraoperative use. The method uses a previously demonstrated approach that combines finite element analysis (FEA) and statistical shape modelling to generate a compact model of prostate gland motion due to insertion of a TRUS probe into the rectum [3,4]. In this study, this technique is used to construct patient-specific, biomechanically-informed statistical motion models (SMMs) from preoperative MR images in order to predict physically realistic deformations as well as providing a well-constrained transformation model for non-rigid registration of MR and TRUS images.

Related previous work reported by Wu et al. [5] and Xu et al. [6] was limited to rigid registration, and is therefore unable to compensate for the significant gland deformation that commonly occurs between MR and TRUS imaging. In the latter study, an initial manual MR-TRUS registration is also required, and quantitative accuracy results are only presented for a phantom [6]. In this work, we propose a "model-to-image" registration approach in which a deformable model of the prostate gland, constructed from segmented MR images prior to a procedure, enables rapid registration to TRUS images acquired during the procedure with minimal user interaction. Although there is a significant time overhead associated with generating such a model, much of the preoperative processing is performed automatically and the level of user interaction required is feasible for the purposes of surgical planning, especially if advanced image segmentation tools are employed.

2 Methods

An overview of the registration method developed in this work is illustrated in Fig. 1. The following two-stage scheme is proposed:

a) *Planning Stage* (before a surgical procedure): i) Build a patient-specific finite element (FE) mesh of the prostate gland and surrounding anatomy from preoperative

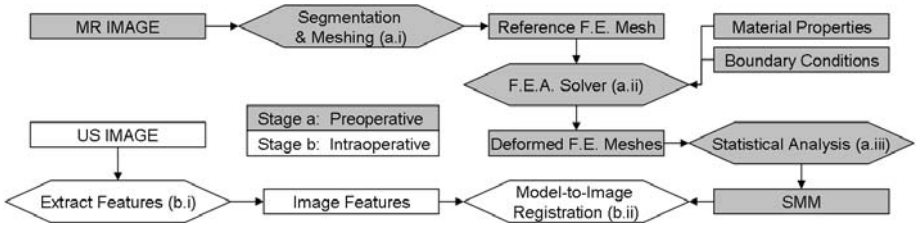


Fig. 1. An overview of the registration method. *Preoperative* processes and data are shaded grey, whereas *intraoperative* processes and data are not shaded.

MR image; ii) Perform a series of FEA simulations of gland motion using randomly sampled material properties and boundary conditions; iii) Construct a SMM by applying principal component analysis (PCA) to the predicted FE mesh node displacements.

b) *Registration Stage* (during a procedure): i) Compute the TRUS image feature vectors by computing second derivatives of the image intensity; ii) iteratively optimise the rigid-body and SMM shape parameters until the likelihood of a particular set of registration parameters given the TRUS image is maximised.

In Section 2, a detailed description of these two stages is provided.

2.1 Statistical Motion Model Generation

For the purposes of this study, diagnostic MR images were manually segmented into regions, which defined the geometry of the prostate gland (divided anatomically into the central and peripheral zones), the pelvis bone, the rectum and the bladder at the base of the prostate (see Fig. 2). The prostate gland was initially described using a spherical harmonic representation, which was then converted into a triangulated surface mesh. The lower part of the pelvis was also meshed. The surfaces were then imported into the commercial FEA software package ANSYS (ANSYS Europe Ltd., Oxfordshire, UK) and a FE model constructed with 50-60,000 tetrahedral elements using the solid modeling tools provided by the software. Ten-node tetrahedral elements were used as these support non-linear geometries using unstructured meshes. The mesh was refined around the region of rectum so that the TRUS probe could be modeled directly in simulations without remeshing.

Elements within all regions of interest were labeled and each assigned material properties randomly sampled from a physiological range. All tissues were assumed to behave as isotropic, linear elastic materials. Since it can been argued that the assumption of incompressibility (Poisson's ratio, $v = 0.5$) may not be appropriate for organs such as the prostate because of gain and loss of blood and other fluids, and the presence of a collapsible urethra, both the Young's modulus and the Poisson's ratio assigned to different materials in the FE model were assumed to be unknown and therefore sampled randomly in the simulations.

Boundary conditions for each FEA simulation were determined as follows: The displacement on the surface of the pelvis was set to zero for all simulations. A random configuration of the TRUS probe in terms of its pose and the diameter of the water-filled sheath were set for each simulation [4].

After assigning sampled material properties and boundary conditions for each of 500 simulations, the node displacements were computed using the preconditioned conjugate gradient iterative equation solver in ANSYS.

Since correspondence was established between the deformed prostate models, PCA was applied directly to the 3D displacements of the mesh nodes, as follows: For each of M (= 500) simulated gland deformations, the displacement of each of N nodes in the prostate gland mesh was calculated and combined to form a $3N\times1$ vector $\mathbf{d}$, which describes the predicted motion of the prostate gland for a particular set of material properties and boundary conditions. The principal modes of variation in $\mathbf{d}$ were then calculated using PCA. If $\mathbf{m}_0$ represents the undeformed gland and is a vector containing the 3D coordinates of the nodes of the original FE model determined from the MR image, then a deformed gland is defined by vector $\mathbf{m}$, given by:

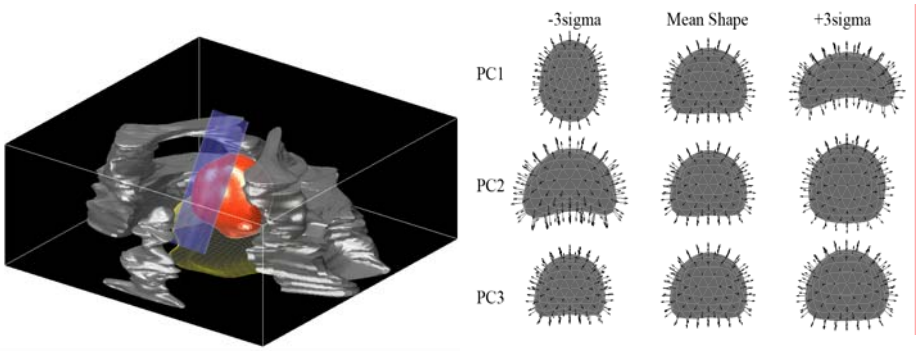


Fig. 2. Left: An illustration of surface meshes derived from an MR image. The TRUS probe (with sheath) is approximated by a cylinder (shown in blue). Right: The first 3 modes (PC1, PC2 & PC3) of an SMM showing the variation in prostate shape. The normal vectors at the nodes of the triangulated surfaces are indicated by arrows.

$$\mathbf{m} = \mathbf{m}_0 + \bar{\mathbf{d}} + \sum_{i=1}^{L} c_i \mathbf{e}_i \tag{1}$$

where $\bar{\mathbf{d}}$ is the mean node displacement vector, $\mathbf{e_i}$ is the i^{th} eigenvector of the covariance matrix, and c_i is a scalar weight. $L \leq M$ was chosen so that the SMM covered >99% of variance in the training data, typically, $L \sim 15$. Additionally, the normal vectors at the nodes (vertices) of the triangulated surface were computed. An example of the shape changes of a prostate model corresponding to the first three modes of the shape variation, are illustrated in Fig. 2.

2.2 Ultrasound Surface Feature Detection

One distinct feature in MR and TRUS images of the prostate gland is the capsule surface. In the proposed registration method, vector representations of this surface, computed independently from the MR-derived model and the 3D TRUS image, are used to drive the model-to-image registration by maximising the similarity between these vectors. In this formulation, the surface of a deformable model, given a set of registration parameters (i.e. rigid-body parameters and shape parameters defined by $\{c_1, c_2, \ldots, c_L\}$), is uniquely defined by the surface normal vector field $\mathbf{u}(\mathbf{x})$, where $\mathbf{x}$ is a position vector that defines the 3D co-ordinates of a point in the model space, and $\mathbf{u}$ is a 3D vector function that defines the surface normal at a particular point. By definition $\mathbf{u}$ is zero at all points not lying on the model surface.

Similarly, a surface normal vector field, denoted by $\mathbf{v}$, can be estimated for the image using a multiscale filtering technique based on second-order Gaussian derivatives. In such approaches, the Hessian is computed at each voxel for a particular scale. The relative magnitudes of the eigenvalues of the Hessian can then be used to classify the local structure, enhancing blob-, tubular- or sheet-like structures [7]. In this work, an extension of the sheet-like enhancement filter proposed by Descoteaux et al. [7] was derived to take into account the non-uniform ultrasound (US) image intensity characteristics found at boundaries due to the variable angle between a boundary surface and the US beam path. This effect is responsible for artefacts where, for example, the

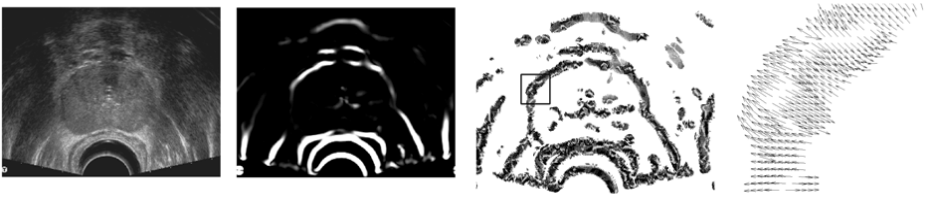


Fig. 3. Example of the surface normal vector field estimated from a 3D TRUS image using the method described in Section 2.2. From left to right: transverse slice through the original TRUS volume; image representing the response of the filter defined in Eq. (3); extracted vector field **v** (projected onto the slice) given by Eq. (4); and a zoomed-in view of a region of interest (shown in the third image) around part of the gland surface.

intensities at the boundary on the lateral sides of the prostate gland are low compared to those on the inferior and superior sides of the gland.

In the original formulation described in [9], the filter response is given by:

$$f_{sheet}(x,y,z) = \exp(-\frac{(R_1)^2}{2\alpha^2})(1-\exp(-\frac{(R_2)^2}{2\beta^2}))(1-\exp(-\frac{(R_3)^2}{2\gamma^2})), \quad (2)$$

where, the ordered eigenvalues λ_1, λ_2 and λ_3 ($|\lambda_1| \leq |\lambda_2| \leq |\lambda_3|$), of the Hessian, were computed at point (x,y,z), $R_1=|\lambda_2/\lambda_3|$, $R_2=|2|\lambda_3|-|\lambda_2|-|\lambda_1||$ and $R_3=(\lambda_1^2+\lambda_2^2+\lambda_3^2)^{0.5}$. For the TRUS data collected in this study, the response of this filter was found to be insensitive to the scalar weighting parameters α, β and γ, and therefore these were set to constant values as suggested in [7]. The width σ of the Gaussian kernel used to compute the Hessian was 1mm in all directions.

If the direction of the US beam is defined by the 3D vector **b**, then the modified filter response is given by:

$$f^*_{sheet} = (\mathbf{n}_3^T \mathbf{b})^2 f_{sheet} \ , \quad (3)$$

where $\mathbf{n}_3(x,y,z)$ is the normalised eigenvector corresponding to the largest eigenvalue (λ_3) of the Hessian, which will be approximately co-linear with the surface normal at the surface. The weighting factor in Eq. (3) reduces the response to noise when direction of the US beam is approximately perpendicular to the surface normal. Finally, the surface normal vector field is given by:

$$\mathbf{v}(x,y,z) = \begin{cases} \mathbf{n}_3, & \text{if } a \leq f^*_{sheet}(x,y,z) \leq b \text{ and } \lambda_3 > 0 \\ \mathbf{0}, & \text{otherwise} \end{cases} \quad (4)$$

where the scalars a and b specify a window in which the filter response is considered to be significant. An example of this vector field is given in Fig.3.

2.3 Definition of a Probability Density Function for the Surface Noise Model

To enable robust model-to-image registration, a similar approach to that described by Staib and Duncan [8] was adopted. The model-to-image registration problem is equivalent to the boundary finding problem considered in [8], and a feature extracted from the image, such as the surface normal vector field, described above, may be considered to be a noise-corrupted version of the surface normal vector field determined from the deformable model **m** (defined in Section 2.1). In this formulation, the probability that a particular image voxel, referenced by the index i in the image space

Ω_{image}, has co-ordinates $\mathbf{y}_i = (x_i, y_i, z_i)$ and an estimated surface normal vector $\mathbf{v}_i$ can be expressed as a probability mixture model as follows:

$$f_N(\mathbf{y}_i, \mathbf{v}_i \mid \mathbf{m}) = \sum_{j \in \Omega_{\text{model}}} h_j f_G(\mathbf{y}_i; \mathbf{x}_j) f_W(\mathbf{v}_i; \mathbf{u}_j) \ , \tag{5}$$

where h_j is a mixing parameter, $\Sigma\, h_j = 1$, j is an index to a discrete point in the model space Ω_{model}, defined by $\mathbf{x}_j = (x_j, y_j, z_j)$, and f_G and f_W are probability density functions that describe Gaussian and bipolar Watson distributions [9] defined as:

$$f_G(\mathbf{y}_i; \mathbf{x}_j) = ((2\pi)^{1.5} \mid \boldsymbol{\Sigma}_j \mid^{0.5})^{-1} \exp(-\tfrac{1}{2}(\mathbf{x}_j - \mathbf{y}_i)^{\mathbf{T}} \boldsymbol{\Sigma}_j^{-1} (\mathbf{x}_j - \mathbf{y}_i)) \tag{6}$$

and

$$f_W(\mathbf{v}_i; \mathbf{u}_j) = C(k) \exp(k(\mathbf{u}_j^{\mathbf{T}} \mathbf{v}_i)^2) = C(k) \exp(k \cos^2 \theta) \ , \tag{7}$$

respectively.

In Eq. (6) a special class of anisotropic Gaussian with two parameters is used where the covariance matrix Σ_j is restricted by a set of orthogonal vectors $\mathbf{w}_d$:

$$\boldsymbol{\Sigma}_j = \sum_{d=1}^{3} \rho_d \mathbf{w}_d \mathbf{w}_d^{\mathbf{T}} \ , \tag{8}$$

where $\mathbf{w}_d$ defines the orientations of the ellipsoid (which defines a surface of constant probability density), and $\mathbf{w}_1$ is set to $\mathbf{u}_j$. The two independent parameters, ρ_1 and ρ_2 ($=\rho_3$) govern the "capture range" in the surface normal direction and in the tangent plane, respectively. For the experiments described here, $\rho_1 = 2\rho_2$.

In Eq. (7) k is a scalar concentration parameter, which is varied depending on the noise level. k was set to a small value $0.1<k<0.5$ in order to weaken the contribution from strong local match. The normalising constant $C(k)$ was estimated by recursive integration to satisfy the requirements of a probability density function. θ is the angle between the model surface normal vector, computed at point j, and the image surface normal vector, computed at voxel i.

2.4 Registration Scheme

The aim of the registration algorithm is to find the optimal registration parameters which maximise the joint probability of the noise. Assuming that the noise values at different voxels are independent [8], we arrive at the following log likelihood objective function:

$$\begin{aligned} f_{obj} &= \log(L(\mathbf{m} \mid \mathbf{I})) = \log \prod_{\Omega_{\text{image}}} P(\mathbf{I} \mid \mathbf{m}) = \log \prod_{i \in \Omega_{\text{image}}} f_N(\mathbf{y}_i, \mathbf{v}_i \mid \mathbf{m}) \\ &= \sum_{i \in \Omega_{\text{image}}} \log \sum_{j \in \Omega_{\text{model}}} h_j f_G(\mathbf{y}_i; \mathbf{x}_j) f_W(\mathbf{v}_i; \mathbf{u}_j) \end{aligned} \tag{9}$$

The expectation maximisation (EM) algorithm provides an efficient means of maximising a likelihood function [10]. An EM algorithm was implemented using Matlab (The Mathworks Inc.), which iteratively optimises the registration parameters in order to maximise Eq. (9).

3 Experiments and Validation Results

All patient data used in this study was obtained from patients recruited to clinical research studies at University College London Hospital (UCLH), approved by the local

Table 1. Target registration errors (TREs) calculated for intra-prostatic landmarks

Case No.		1	2	3	4	5	6	7	All
Number of Landmarks		5	3	3	4	4	4	3	26
TRE (mm)	RMS	1.92	3.67	3.14	1.86	1.57	3.23	3.12	2.66
	Mean	1.84	3.01	3.02	1.07	1.47	3.21	2.91	2.36
	Std.	0.63	2.58	1.04	0.94	0.64	0.46	1.38	1.24
	Max.	2.53	5.98	4.19	2.93	2.36	3.56	4.46	5.98

research ethics committee. All patients gave written consent to participate. Data from 7 patients with prostate cancer were used to validate the method described here. T2-weighted MR image volumes of the prostate gland were acquired prior to template-guided transperineal needle biopsy under general anaesthesia. Immediately before needle insertion, 3D TRUS images of the gland were acquired using a B-K ProFocus scanner (B-K Medical Ltd., Berkshire, UK). A set of parallel 2D transverse B-mode images were captured at 2mm intervals and stored on the scanner. A mechanical stepping device (Tayman Medical Inc., MO, USA) was used to translate the US probe (B-K 8658T, 5-7.5MHz transducer) perpendicular to the axial plane along the rectum.

Each US image was first resampled into a volume with an isotropic voxel dimension of 1mm. At each voxel, the Hessian was computed in the frequency domain using an implementation based on the fast Fourier transform. A quick and simple procedure was used to initialise the pose of SMM with respect to the TRUS volume where two points at the apex and base of the gland were manually identified. Once registered, a dense displacement field (DDF) was computed across the volume of interest by interpolating the final instance of the SMM with a solid FE mesh using a shape function for tetrahedral elements.

Corresponding anatomical landmarks, including cysts, calcifications, the urethra, the puboprostatic ligament, and the junction between the seminal vesicles, the vas deferens and the midline of the gland, were identified manually in both the MR and TRUS volumes. The 3D co-ordinates of landmarks defined in the MR image were then propagated into TRUS co-ordinates using the DDF. For each pair of identified and propagated landmarks, a target registration error (TRE) was calculated, defined as the distance between the manually defined and propagated landmarks in the TRUS space. The MR images were also warped using the DDF to allow a visual assessment of the registration.

The landmark-based results are given in Table 1. The root-mean-square (RMS) TRE over all 7 cases (26 landmarks) was 2.66mm. Fig. 4 illustrates the warped MR images and target TRUS images computed from an example registration.

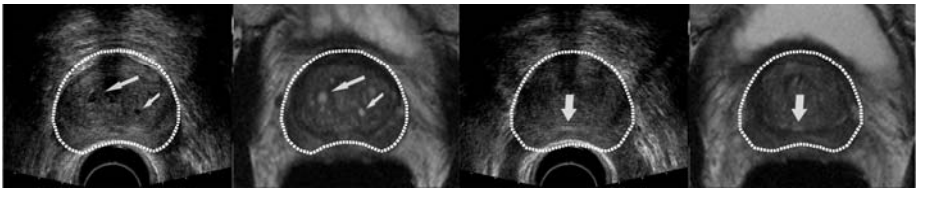


Fig. 4. Example transverse image slices (1st and 3rd images) through a TRUS volume for Case 1 shown with the corresponding warped MR images (2nd and 4th images) following deformable registration. The arrows indicate landmarks which were well aligned.

4 Discussion

A model-to-image registration method is proposed in this study, which uses a combined statistical-biomechanical model built from an MR image. Building the model requires manual segmentation of the MR image and is computationally intensive, but is performed preoperatively, so does not significantly impact the intraoperative workflow. A model to image registration can currently be performed within 2 minutes using a desktop PC with a 2.33GHz Intel® Core™ dual CPU processor and 3GB of RAM. The method achieved sufficiently high accuracy to be clinically useful for MR-targeted prostate biopsy and interventions. Although only the gland surface is registered in this scheme, the use of a deformable finite-element model enables the displacement of internal structures to be rapidly computed.

Acknowledgments. This work is funded by Royal Academy of Engineering/EPSRC Research Fellowship and UCLH/UCL Comprehensive Biomedical Research Centre. The authors would like to thank Tim Carter, Gerard Ridgeway and Xiahai Zhuang for helpful ideas and discussions.

References

1. Ahmed, H.U., Emberton, M.: Active Surveillance and Radical Therapy in Prostate Cancer: Can Focal Therapy Offer the Middle Way? World J. Urol. 26, 457–467 (2008)
2. Kirkham, A.P.S., Emberton, M., Allen, C.: How good is MRI at detecting and characterising cancer within the prostate? Eur. Urol. 50, 1163–1174 (2006)
3. Mohamed, A., Davatzikos, C., Taylor, R.: A combined statistical and biomechanical model for estimation of intra-operative prostate deformation. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2489, pp. 452–460. Springer, Heidelberg (2002)
4. Hu, Y., Morgan, D., Ahmed, H.U., Pendsé, D., Sahu, M., Allen, C., Emberton, M., Hawkes, D., Barratt, D.: Modelling Prostate Gland Motion for Image-guided Interventions. In: Bello, F., Edwards, E. (eds.) ISBMS 2008. LNCS, vol. 5104, pp. 79–88. Springer, Heidelberg (2008)
5. Wu, R., Ling, K.V., Shao, W., Ng, W.-S.: Registration of Organ Surface with Intraoperative 3D Ultrasound Image Using Genetic Algorithm. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 383–390. Springer, Heidelberg (2003)
6. Xu, S., Kruecker, J., Turkbey, B., Glossop, N., Singh, A.K., Choyke, P., Pinto, P., Wood, B.J.: Real-time MRI-TRUS fusion for guidance of targeted prostate biopsies. Comput. Aided Surg. 13(5), 255–264 (2008)
7. Descoteaux, M., Audette, M.A., Chinzei, K., Siddiqi, K.: Bone Enhancement Filtering: Application to Sinus Bone Segmentation and Simulation of Pituitary Surgery. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 247–255. Springer, Heidelberg (2005)
8. Staib, L.H., Duncan, J.S.: Boundary Finding with Parametrically Deformable Models. IEEE Trans. PAMI 14, 1061–1075 (1992)
9. Watson, G.S.: Statistics on Spheres. Wiley, New York (1983)
10. Figueiredo, A., Gomes, P.: Performance of the EM Algorithm on the Identification of a Mixture of Watson Distributions Defined on the Hypersphere. Statistical Journal 4(2), 111–130 (2006)

Combining Multiple True 3D Ultrasound Image Volumes through Re-registration and Rasterization

Songbai Ji[1], David W. Roberts[2,3], Alex Hartov[1,2], and Keith D. Paulsen[1,2]

[1] Dartmouth College, Thayer School of Engineering, HB8000, Hanover, NH 03755
[2] Norris Cotton Cancer Center, Lebanon, NH 03756
[3] Dartmouth Hitchcock Medical Center, Lebanon, NH 03756
{songbai.ji,david.w.roberts,alex.hartov,
keith.d.paulsen}@dartmouth.edu

Abstract. We present an accurate and efficient technique to combine and rasterize multiple 3D ultrasound (3DUS) image volumes originally presented in spherical coordinates into a single, 3D Cartesian image that uniformly samples the total field of view. To ensure the consistency of merged image content in overlapping regions, image re-registration was performed by maximizing mutual information (MI). The technique was applied to 22 3DUS image volumes obtained during five neurosurgical patient cases. The computational cost of the approach increases linearly with the number of images involved (average time to combine and rasterize one pair of 3DUS images was 1.5 sec). Interpolation was approximately 20% more accurate in overlapping regions when re-registration was performed before rasterization and minimized feature loss and/or blurring that was evident without re-registration. In addition, we report the average translational (35.2 mm) and rotational (38.5^{o}) capture ranges for the MI re-registration of two volumetric 3DUS images. The technique is applicable in any clinical application in which volumetric true 3DUS is acquired.

1 Introduction

Ultrasonography is an important imaging technique with a wide range of both diagnostic and intraoperative applications. Conventional 2D ultrasound (2DUS) is currently the most commonly used imaging scheme where multiple freehand sweeps are acquired to sample the region of interest. Recently, we have integrated volumetric true 3D ultrasound (3DUS) into image-guided neurosurgery, during which image volumes are generated from a dedicated ultrasound transducer without the need for freehand sweeps or 3D image reconstruction [1]. However, the 3DUS image space is represented in an unconventional spherical coordinate system, making it incompatible with most processing and visualization software that expects Cartesian coordinates. In addition, multiple acquisitions, as opposed to a single snapshot, are still often recorded (e.g., to increase the overall sampling of the region of interest). Therefore, a technique that combines multiple 3DUS image volumes into one Cartesian 3D dataset is needed to enable application of existing software as well as to improve the efficiency of any subsequent image processing of the 3DUS acquisitions.

This paper presents an accurate and efficient technique to combine arbitrarily-oriented 3DUS volumes into a single Cartesian coordinate system. The consistency of

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 795–802, 2009.

the combined image content in overlapping regions is ensured through re-registration with mutual information (MI), the importance of which is demonstrated by comparing rasterized images generated with and without the re-registration. We show that image features can become lost or significantly blurred without re-registration. In addition, the capture ranges of two true 3DUS image volumes registered with MI are quantitatively evaluated and found to be much larger than those of tracked 2DUS acquisitions.

2 Material and Methods

Five patients (3 male, 2 female; average age of 48) undergoing open cranial brain tumor resections (three low grade gliomas, one high grade glioma and one meningioma) with deployment of volumetric 3DUS were included in the study. A set of volumetric 3DUS images was acquired (3–9 image volumes for each patient and 22 in total; time interval between two consecutive volumes was approximately 10–15 sec) before dural opening using a dedicated transducer (X3-1 broad band matrix array) and ultrasound system (iU22, Philips Healthcare, N.A.; Bothell, WA). All image acquisitions were configured to cover the maximum angular ranges allowed [1]. The scan-depth was set to 140–160 mm to capture the parenchymal surface contralateral to the craniotomy.

2.1 Geometric Transformation and Interpolation for Volumetric 3DUS

Interpolation is essential for 3DUS image rasterization. We have recently developed an accurate and efficient trilinear interpolation scheme out of necessity because the vendor of the 3DUS scanner did not openly provide conversion functionality to transform scans into Cartesian coordinates. The interpolation was achieved by first converting physical points into an integer Cartesian grid space, similarly to that reported in [2]. Briefly, the dimensions of the 3DUS image matrix and the ranges in depth (r, in mm), and lateral (θ; in degrees) and medial angles (φ; in degrees) determine the step sizes in each direction (Fig. 1a). Therefore, the indices of the row, column and slice of a voxel in a typical 3DUS image determine its physical location. Conversely, for any given point in the physical space (Fig. 1b), its equivalent coordinates in grid space (i, j, k; *not* necessarily integers) can be uniquely determined (Fig. 1c; [1]).

The intensity value at p or its equivalence, p', can then be linearly interpolated using the standard finite element trilinear shape functions [3]. Specifically, the 8 neighboring voxels formulate an 8-node hexahedral element in grid space, which is further transformed into natural coordinates (Fig. 1d). The trilinear shape functions for a normalized hexahedral element are expressed as:

$$N_a = \frac{1}{8}(\xi + \xi_a)(\eta + \eta_a)(\zeta + \zeta_a), \tag{1}$$

where a indexes from 1 to 8, representing the eight neighboring nodes in the element, while ξ, η, and ζ are three normalized coordinates (subscript indicates value at node a). The intensity at p is, therefore, calculated as the weighted sum of the intensities at the neighboring voxels according to the following equation (Fig. 1d):

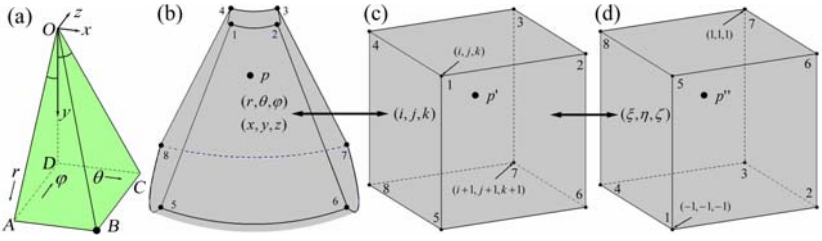


Fig. 1. (a) A typical 3DUS voxel (B) in physical space. (b–d) Sequential transformations of a typical point from physical space (p) to grid space (p'), and subsequently to natural coordinates (p'') of a hexahedral element determined by the eight surrounding voxels.

$$I(p) = I(p') = I(p'') = \sum_{a=1}^{8} N_a \times I(a) \ , \tag{2}$$

where $I(a)$ is the intensity value at the corresponding voxel. The accuracy of the trilinear interpolation algorithm has been verified with clinical 3DUS images, and is an improvement over both voxel nearest neighbor or distance weighting algorithms [1].

2.2 Combination of Multiple Images

In order to combine multiple 3DUS images, a common coordinate system is required, which was accomplished through optical tracking (Northern Digital Inc., London ON, Canada) that continuously monitored two infra-red light emitting sources rigidly coupled with the patient's head (patient tracker) and the US scan-head (US tracker), respectively. Image transformation is illustrated in Fig. 2a, in which ${}^{tracker}T_{US}$ was obtained by scan-head calibration (accuracy of approximately 2 mm; [4]). An arbitrary 3DUS image was transformed into a pre-selected 3DUS volume (chosen as the first 3DUS image acquired for each patient) coordinate system by:

$${}^{US1}T_{US2} = \mathrm{inv}\left({}^{tracker}T_{US}\right) \times \mathrm{inv}\left({}^{patient}T^{(1)}_{tracker}\right) \times {}^{patient}T^{(2)}_{tracker} \times {}^{tracker}T_{US}, \tag{3}$$

The accuracy of the image transformation depended on the accuracies of the Polaris tracker (error <1 mm) as well as the US scan-head calibration (error of 2–3 mm; [4]). To improve the accuracy of image transformation, an inter-image re-registration was performed to maximize image alignment (see 2.3 for details) before rasterization.

After all 3DUS images were transformed into a common coordinate system (the relative size of the overlapping region between two image volumes with the same scan-depth was 79.9% on average (ranged from 56.7% to 94.0%)), a 3D bounding box was established with its major axes parallel to those of the first 3DUS image acquisition (Fig. 2b). A set of regularly spaced voxels was generated to fill the bounding box. The spacing between voxels determined the voxel size, and was chosen to be 1.0 mm along all three directions. The image intensities of these voxels were determined through interpolation (see section 2.1 for details). In addition, an extra band of 5 zero intensity pixels was padded along the boundaries of the bounding box to ensure complete sampling of the combined imaging volume.

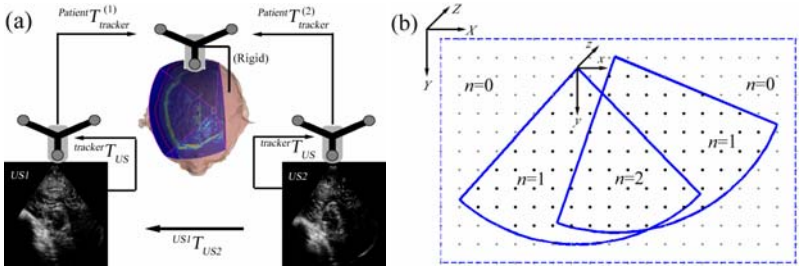


Fig. 2. (a) Illustration of image coordinates involved in transforming 3DUS images. (b) Illustration of image combination and rasterization. Image intensity values of voxels enclosed by multiple 3DUS images ($n \geq 1$) were obtained by averaging.

Apparently, each Cartesian voxel may be physically enclosed by any number of the 3DUS acquisitions (n = 0, or ≥1; Fig. 2b). When n=0 (i.e., the voxel was not enclosed by any 3DUS image), a zero-intensity was assigned. Otherwise (i.e., $n \geq 1$), an averaging scheme was used to prescribe a unique intensity value by interpolating across all of the 3DUS images in which the voxel was enclosed.

2.3 Inter-image Re-registration through Maximization of MI

Inter-image re-registration through MI (Insight Segmentation and Registration Toolkit; ITK version 3.8 [5]) was used to correct errors in the transformations (i.e., Eqn. 3) required to place all 3DUS volumes into a common coordinate system. The first 3DUS acquisition was chosen as the fixed image while all of the rest of the volumes were treated as moving images. In total, (N–1) re-registrations were performed, where N is the number of 3DUS acquisitions for a particular patient.

Gaussian smoothing (kernel of 5×5) of both the fixed and moving images as well as thresholding of the moving images were performed to improve the robustness of the registration. The initial transformation obtained from the tracking system (Eqn. 3) served as the starting point for re-registration with Mattes version of MI [5] as the image similarity measure. Multithreading was enabled and a steepest gradient descent optimization was employed to maximize the MI. Convergence was reached when either the net change in MI was less than 10^{-3} or the number of iterations reached a preset maximum value of 200 [5]. With image re-registration, the adjusted version of Eqn. 3, which transformed an arbitrary 3DUS volume into the coordinate system of the first 3DUS acquisition can be written as:

$$^{US1}T_{US2}^{adjust} = T_{adjust} \times \mathrm{inv}\left(^{tracker}T_{US}\right) \times \mathrm{inv}\left(^{patient}T_{tracker}^{(1)}\right) \times {}^{patient}T_{tracker}^{(2)} \times {}^{tracker}T_{US}, \quad (4)$$

2.4 Capture Range

To quantify the capture range of registration between two volumetric 3DUS acquisitions, the first two 3DUS volumes recorded in each patient case were selected, rasterized, and then registered. The converged locations of all non-zero intensity voxels in the thresholded moving image (defined as the "true" locations) were subsequently

obtained. The centroid of these voxels was defined as the origin of a local coordinate system (O_{local}). The moving image was then transformed away from the "true" locations over a specified range (0–60 mm in translation and 0–60° in rotation about O_{local}, respectively). For each patient, a total of 400 translational and 400 rotational perturbations were generated, in which the magnitudes of the translational and rotational perturbations linearly increased over the specified range. In addition, the directionality of the translation as well as that of the rotational axis passing through O_{local} was randomly selected from a uniform distribution. The distances from the locations of converged voxels with respect to their corresponding "true" positions were calculated, and the average (distance) error was determined. Successful registration was defined to occur when the average distance error was less than 2 mm. Successful registrations were counted and the capture range was defined as the largest misalignment at or below which the registration success rate was at least 95% across all of the patient cases.

2.5 Data Analysis

To demonstrate the importance of inter-image re-registration, representative images were qualitatively compared with those obtained when no image re-registration was otherwise performed. We also demonstrated the differences by transforming the second 3DUS volume into the coordinates of the first with and without re-registration to interpolate intensities at locations defined by the transformed voxels in the second image based on their corresponding intensities in the first. Absolute differences in image intensities between the interpolated and the "ground-truth" second images indicate the influence of image re-registration on the interpolation accuracy. Scatter plots of the average distance error relative to the initial translational and rotational misalignments as well as success rate curves were generated.

Image rasterization was implemented in C with multithreading enabled by OpenMP [6], and was compiled in Matlab (Matlab R2008b; The Mathworks, Natick, MA). Image rasterization and registration were executed on an 8-CPU Intel Xeon computer running Ubuntu Linux 6.10 (2.33GHz, 8G RAM). All data analyses were performed in Matlab. We report (i) the typical computational cost of registering one pair of 3DUS volumes and combining and rasterizing multiple 3DUS acquisitions, (ii) the interpolation accuracy in overlapping regions with and without image re-registration, and (iii) the capture ranges of the intra-modality image registration.

3 Results

Seventeen (17) image re-registrations were performed and they all converged within 100 iterations (typically 30–70 iterations). The average computational cost for each re-registration was 31 sec. In addition, the amount of transformation (translation and rotation) required to align the moving images with respect to the fixed image was 2.0±0.5 mm (range of 1.3–3.5 mm). Image transformations were adjusted by Eqn. 4.

Subsequently, image rasterization was performed to uniformly sample the combined image volume. Clearly, the computational cost of image combination and rasterization depended linearly on the number of 3DUS volumes involved. The average cost to combine and rasterize each 3DUS acquisition was 1.5 sec. Representative

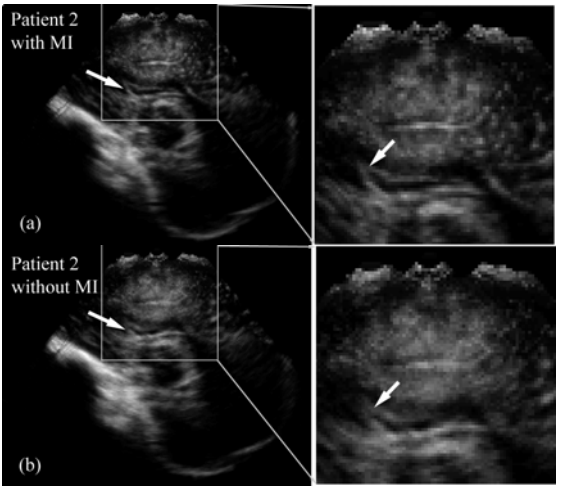


Fig. 3. Combined and rasterized 3DUS images (a) with and (b) without inter-image re-registration for a representative patient. The combined image is significantly sharper when re-registration was applied before rasterization (see arrows and enlarged view of image inset).

2D cross-sectional images passing through the center of the acquisition volume are shown for a representative patient (Fig. 3a). For comparison, the 3DUS images were also combined and rasterized *without* applying the re-registrations, and the corresponding results are also shown (Fig. 3b). Apparently, the combined 3DUS image was sharper (and less fuzzy in appearance) when image re-registration was applied. This is not surprising because tissues in the overlapping regions generally presented the same image intensities (given that the angular changes of the transducer were usually small due to the confinements of the craniotomy). Therefore, image blurring would certainly occur when misalignment was present, demonstrating the importance of image re-registration to minimize any misalignment before rasterization.

The significance of re-registration between 3DUS volumes was apparent when overlaying two image acquisitions in the same space, where clearly misalignment of features between image acquisitions was significantly minimized with the application of registration (Fig. 4). The significance was further demonstrated by transforming the second 3DUS image into the coordinates of the first and comparing the absolute differences in image intensities between the interpolated and the "ground-truth" second image. With re-registration, the absolute difference was 9.5±2.8 for five patient cases, whereas it was 11.7±4.1 otherwise (all images were 8-bit grayscale).

Scatter plots of the average distance error relative to the initial translational and rotational misalignments when perturbed away from the "true" locations are shown in Fig. 5ab. With the results pooled across patients, the registration success rate was plotted against the initial misalignments (Fig. 5cd). The intersection of the success rate curve with the horizontal dashed line at 95% indicates that the overall translational and rotational capture range was 35.2 mm and 32.8^{o}, respectively. As a comparison, when the threshold for average distance error of 4 mm was used (instead of 2 mm) to define a successful registration, the translational and rotational capture ranges differed by less than 2 mm and 1^{o}, respectively. When a success rate of 90% was considered sufficient, the capture ranges were improved to 39.2 mm and 38.9^{o}.

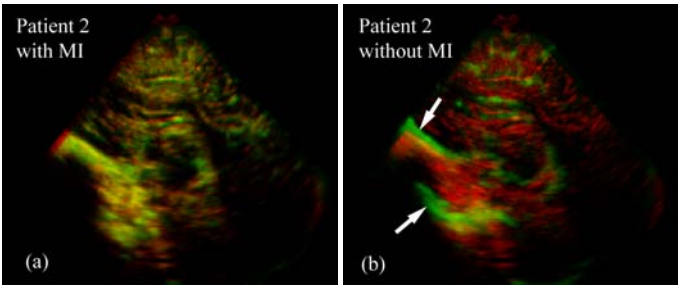


Fig. 4. Significance of registration is apparent when comparing representative overlays of two 3D volumes in the same space generated (a) with or (b) without registration. Arrows in (b) indicate areas of significant misalignment (approx. 3.5 mm) corrected by MI re-registration.

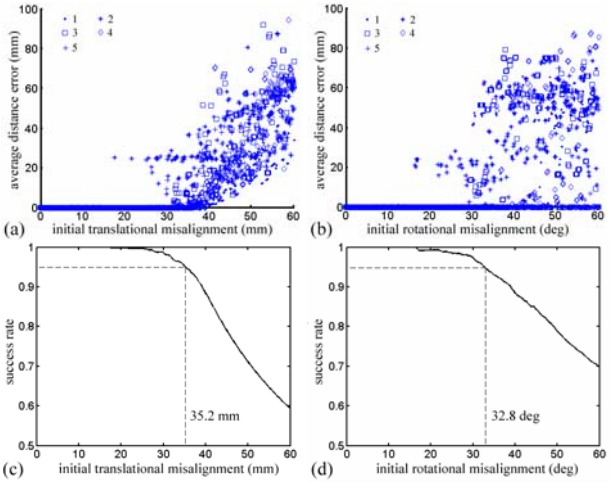


Fig. 5. Scatter plots of average distance error vs. initial (a) translational and (b) rotational misalignment, along with the corresponding (c) translational and (d) rotational success rate curves

4 Discussion and Conclusion

We have presented an accurate and efficient image combination and rasterization technique for generating a single Cartesian 3D image volume from multiple 3DUS image acquisitions that has been used in five neurosurgical patient cases. The computational cost of the approach increased linearly with the number of images involved with an average of 1.5 sec for combining and rasterizing each 3DUS acquisition. Re-registering 3DUS images before rasterization significantly increased the consistency of image content in overlapping regions. The average magnitude of transformation required to re-align the 3DUS volumes was 2.0 mm. The importance of re-registration in producing sharper images in overlapping regions was demonstrated by visually comparing the results generated *without* re-registration. The enlarged views in Fig. 3 clearly indicate that features becomes lost or significantly blurred without the

MI re-registration. By comparing the absolute differences between the interpolated and "ground-truth" image intensities, the interpolation was found to be approximately 20% more accurate in overlapping regions when image re-registration was performed. The residual interpolation error (9.5±2.8) was likely due to inherent noise in the 3DUS acquisitions (e.g., caused by varying acoustic coupling to brain parenchyma).

The capture ranges of intra-modality registration between two volumetric 3DUS acquisitions (35.2 mm and 35.8^{o}) are similar to those in reports of reconstructed 3DUS images in the literature (25.5 mm [7] and 40^{o} [8]) and much larger than those we have obtained when re-registering tracked 2DUS. Given the large capture ranges, it may be possible to register two 3DUS images without the need for optical tracking if sufficient overlap is present, which may further simplify the technique.

Although the approach presented in this study was demonstrated with neurosurgical cases, it is also applicable in other imaging contexts (e.g., imaging the abdomen, pelvis, etc.) in which volumetric true 3DUS is deployed. As long as the images are tracked or they have sufficient overlap to allow accurate inter-image transformation, multiple volumes can be combined and rasterized into a single 3D Cartesian coordinate system. Because only one image volume (instead of multiple volumes) is produced, the efficiency of subsequent image processing can be significantly improved.

Acknowledgment. Funding from the National Institutes of Health grant R01 EB002082-11 and support from Philips Medical Systems for the iU22 3D ultrasound system are acknowledged**.**

References

1. Ji, S., Hartov, A., Fontaine, K., Borsic, A., Roberts, D.W., Paulsen, K.D.: Coregistered volumetric true 3D ultrasonography in image-guided neurosurgery. In: Miga, M.I., Cleary, K.R. (eds.) Proceedings of SPIE, SPIE, Bellingham, WA, vol. 6918, p. 69180F (2008)
2. Duan, Q., Angelini, E., Song, T., Laine, A.: Fast interpolation algorithms for real-time three-dimensional cardiac ultrasound. In: Proceedings of the 25th Annual International Conference of the IEEE EMBS, pp. 1192–1195 (2003)
3. Zienkiewicz, O.C., Taylor, R.L., Zhu, J.Z.: The finite element method: its basis and fundamentals, 6th edn. Elsevier Butterworth-Heinemann, Amsterdam (2005)
4. Pallatroni, H., Hartov, A., McInerney, J., et al.: Coregistered ultrasound as a neurosurgical guide. Stereotactic and functional neurosurgery 73, 1–4 (1999)
5. Insight Segmentation and Registration Toolkit (ITK), `http://www.itk.org/`
6. OpenMP, `http://openmp.org/wp/`
7. Shekhar, R., Zagrodsky, V.: Mutual information-based rigid and nonrigid registration of ultrasound volumes. IEEE Tran. Med. Imag. 21(1), 9–22 (2002)
8. Slomka, P.J., Mandel, J., Downey, D., Fenster, A.: Evaluation of voxel-based registration of 3-D power Doppler ultrasound and 3-D magnetic resonance angiographic images of carotid arteries. Ultrasound in Med. & Biol. 27(7), 945–955 (2001)

Biomechanically Constrained Groupwise US to CT Registration of the Lumbar Spine*

Sean Gill[1], Parvin Mousavi[1], Gabor Fichtinger[1], Elvis Chen[1], Jonathan Boisvert[1,2], David Pichora[1], and Purang Abolmaesumi[1]

[1] Queen's University, Kingston, ON, Canada
[2] Institute for Information Technology, National Research Council Canada, Ottawa, ON, Canada
pmousavi@cs.queensu.ca

Abstract. Registration of intraoperative ultrasound (US) with preoperative computed tomography (CT) data for interventional guidance is a subject of immense interest, particularly for percutaneous spinal injections. We propose a biomechanically constrained group-wise registration of US to CT images of the lumbar spine. Each vertebra in CT is treated as a sub-volume and transformed individually. The sub-volumes are then reconstructed into a single volume. The algorithm simulates an US image from the CT data at each iteration of the registration. This simulated US image is used to calculate an intensity based similarity metric with the real US image. A biomechanical model is used to constrain the displacement of the vertebrae relative to one another. Covariance Matrix Adaption - Evolution Strategy (CMA-ES) is utilized as the optimization strategy. Validation is performed on CT and US images from a phantom designed to preserve realistic curvatures of the spine. The technique is able to register initial misalignments of up to 20mm with a success rate of 82%, and those of up to 10mm with a success rate of 98.6%.

1 Introduction

Spinal injections for back-pain management are carried out on a frequent basis in hospitals and radiological clinics. Currently, these procedures are performed under fluoroscopy or CT guidance in specialized interventional radiology facilities, and thus incur a major financial burden on the healthcare system. Another drawback with the current practice is patient and surgeon exposure to X-ray radiation. The goal of this research is to design a spine intervention system that uses US for guidance. This would greatly reduce the exposure of both the patient and the physician to ionizing radiation and allow the procedure to be performed outside of a specialized facility. The use of only US for guidance has its own difficulties. In particular, due to the significant level of occlusion in spinal US images, it can be difficult to accurately identify the appropriate injection site. For this reason, in this paper, we consider the fusion of intraoperative US with preoperative CT as a means of guidance for spinal injections (Figure 1).

* This work has been partially funded by NSERC and CIHR. Special thanks to A. Lang, K. Wang, S. Lyman, J. Lazazzera and M. Kunz.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 803–810, 2009.

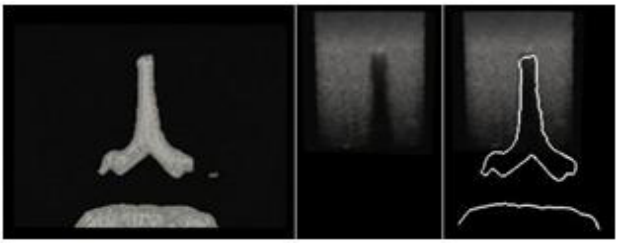


Fig. 1. Axial slice from a CT volume of the spine phantom (left), corresponding US slice (center) and an overlay of the CT bone contours with the corresponding US slice (right)

Point- and surface-based registration of US to CT data often require manual intervention and segmentation of US data, which is time consuming and susceptible to errors. To avoid these problems, we chose to focus on automated intensity-based registration methods. Among previously proposed intensity based methods for registration of US to CT data, Winter et al. [1] propose to define the bone surface in CT that is visible in an US image, with the similarity calculated as the US pixel intensities overlapping the surface. This requires a priori knowledge of the direction and orientation of the US probe. Penney et al. [2] present a method where voxel values in the US and CT data are converted to the probability of representing a bone edge. In order to have clinically relevant probabilities, a large set of prior CT and US images would have to be manually segmented.

In Wein et al. [3,4], density information from CT data is used to iteratively simulate an US image throughout the registration process, thereby optimizing the simulation as the registration proceeds. This has the benefit of not requiring any previous knowledge of the orientation of the US probe. Although the simulation is a simplified take on the physics of ultrasound beam propagation, it is sufficient for registration purposes. This algorithm is extended in Shams et al. [5] to create a more realistic simulation for training physicians and technicians in the use of US imaging. The technique requires preprocessing to create a scatter volume of the CT data using the Field II simulator, which remains time consuming. In Reichl et al. [6] and Kutter et al. [7] the US simulation and registration is implemented in GPU, resulting in a dramatically decreased algorithmic run time. Gill et al. [8] propose an extension of the work from Wein et al. [3,4], a groupwise US to CT registration of vertebrae L3 to L5. The algorithm allows free motion of the vertebrae and registers all three simultaneously. The drawback of this approach is that free motion of vertebrae during registration can lead to biologically unrealistic alignments.

Here we propose an algorithm that extends the groupwise registration presented in Gill et al. [8]. The algorithm allows independent motion of the vertebrae, but constrains their motion based on a well known biomechanical model. The registration is tested on a phantom printed using a surface model of a patient's spine which preserves a clinically realistic curvature of the spine. We present the results of the registration with various weights for the fusion of the biomechanical model with the intensity-based similarity metric.

2 US to CT Registration

The registration workflow can be seen in Figure 2. The CT volume is initially cut into sub-volumes, each containing a single vertebra. Voxels in the sub-volumes corresponding to bone from an external vertebra are masked. The registration treats each vertebra independently, allowing for six degrees of freedom. All vertebrae are registered simultaneously, resulting in an optimization with n×6 parameters, where n is the number of vertebrae being registered. After the transformations are applied, the sub-volumes are reconstructed into a single volume. For any overlapping voxels the maximum intensity is selected for the reconstructed volume, thus preserving bone structure. Any gaps not in the final volume are filled with a default value approximating the intensity of soft tissue in CT. The US simulation is applied to this reconstructed volume. CMA-ES is used as the optimization strategy, as Gill et al. [8] found it to be robust for US to CT simulation and registration.

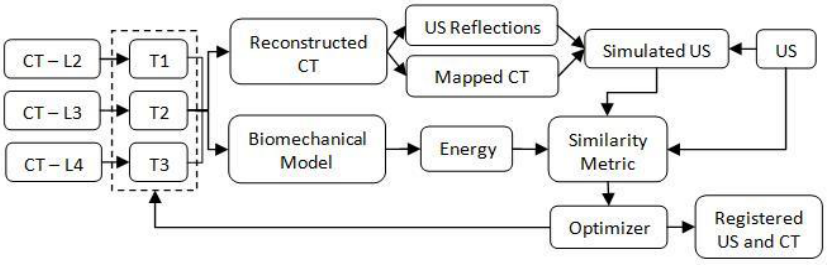


Fig. 2. Workflow of the biomechanically constrained groupwise US to CT registration

There are three distinct steps in the simulation of US from CT: The Simulation of the US reflection from CT, the mapping of the CT values to those found in US, and calculations of the weights for these to images and a bias. The weights are chosen so that the simulation best represents the real US image.

The simulated ultrasound reflections model the ultrasound beam passing through the tissue as a ray. The assumption is made that the CT intensities (in Hounsfield units) can be related to the acoustic impedance values used to calculate ultrasound transmission and reflection. The simulated beam passes through each column of the volume. The transmission and reflection of the beam is calculated at each voxel based of the following equations,

$$\Delta r(x, y, d) = (d^T \nabla \mu(x, y)) \frac{|\nabla \mu(x, y)|}{(2\mu(x, y))}, \tag{1}$$

$$\Delta t(x, y) = (1 - \left(\frac{|\nabla \mu(x, y)|}{(2\mu(x, y))} \right)^2, \tag{2}$$

$$r(x, y) = I(x, y - 1) \Delta r(x, y, d), \tag{3}$$

$$I(x,y) = \begin{cases} I(x,y-1)\Delta t(x,y), & |\nabla\mu(x,y)| < \tau \\ 0, & |\nabla\mu(x,y)| \geq \tau \end{cases}, \tag{4}$$

where d is the direction of the US beam, μ is the intensity of the CT image, ΔR is reflection coefficient, r is the simulated reflection intensity, Δt is the transmission coefficient, τ is the threshold for full reflection and I is the intensity of our simulated US beam. Any gradient value greater than a set threshold (450 h.u. in our simulations) causes full reflection of the US beam intensity at that point, setting the incoming US beam intensity for all subsequent points on the scan line to zero. A log-compression is applied to the simulated reflection image to amplify small reflections,

$$r(x,y) = \frac{\log(1 + ar(x,y))}{\log(1+a)}. \tag{5}$$

The CT intensities are mapped to values closer to those corresponding to the tissues in the US data using an approximation of the curve presented in Wein et al. [3,4],

$$p(x,y) = 1.36\mu(x,y) - 1429. \tag{6}$$

The final step of the US simulation is the weighting of the simulated US reflection, the mapped CT and a bias term. A least-squares optimization is used to calculate the weights, such that the values in the simulation best match the corresponding intensities in the real US volume. The final simulation is calculated as,

$$f(x,y) = \begin{cases} \alpha p(x,y) + \beta r(x,y) + \gamma, & I(x,y) > 0 \\ 0, & I(x,y) = 0 \end{cases}, \tag{7}$$

where f is the simulated US image and α, β, γ are the weights for their respective images. We do not include any voxels that are occluded in the simulation as part of the weight calculation. All occluded voxels in the US simulation are set to zero. Occluded voxels are identified as any voxel where the intensity of the incoming simulated US beam is zero. Similarity between the actual US image and simulated US image is calculated using the Linear Correlation of Linear Combination (LC^2) metric presented by Wein et al. [3,4],

$$LC^2 = \frac{\sum(U(x,y) - f(x,y))^2}{N \times Var(U)}, \tag{8}$$

where N is the number of overlapping voxels between the US and CT images, and U is the actual ultrasound image intensity. All voxels, including occluded voxels, are used in the calculation of the similarity metric.

3 Biomechanical Model

Allowing free motion of vertebrae during registration is not ideal as it does not distinguish between anatomically realistic orientations of vertebrae and orientations where the vertebrae are colliding, unreasonably oriented or far apart. We

propose the use of a biomechanical model of the spine to constrain the registration and to favour anatomically acceptable alignments. The biomechanical model we use models the relation between the displacement of the intervertebral structures and the reaction forces and moments[9] :

$$K = \begin{bmatrix} 100 & 0 & 50 & 0 & -1640 & 0 \\ 0 & 110 & 0 & 150 & 0 & 580 \\ 50 & 0 & 780 & 0 & -760 & 0 \\ 0 & 150 & 0 & 1.48E5 & 0 & -8040 \\ -1640 & 0 & -760 & 0 & 1.52E5 & 0 \\ 0 & 580 & 0 & -8040 & 0 & 1.53E5 \end{bmatrix} \left[\text{Nmm rad}^{-}1\right], \tag{9}$$

where K is the stiffness matrix representing the intervertebral structures. This stiffness matrix is multiplied with a vector x representing the change in translation and rotation of the intervertebral link. For our case, x is calculated as the relative transform between two consecutive vertebrae. Each vertebra is expected to have no change in rotational orientation and no translation along the x and y axes. The expected translation along the coronal axis is defined as the mean distance between the centers of consecutive vertebrae in the patient's CT data. Note that this is meant as an approximation of the vertebral resting position.

The energy of the system based on the relative transformations between the vertebrae is calculated using the general spring equation,

$$U = \frac{1}{2}(x^T K x). \tag{10}$$

The energy is calculated across all vertebrae and normalized based on the energy of a maximum misalignment (15 mm translation along each axis and 15 rotation about each axis),

$$E = \frac{(U_{\text{L2,L3}} + U_{\text{L3,L4}})}{2 \times U_{\max}}, \tag{11}$$

where E is the normalized energy of the system, $U_{\text{L2,L3}}$ and $U_{\text{L3,L4}}$ represent the energy of the model calculated from the relative transforms between L2-L3 and L3-L4, respectively, and $U_{\max}$ is the energy of the maximum misalignment.

This normalized energy is then combined with the LC^2 metric to give the Biomechanically Constrained Linear Correlation of Linear Combination ($BCLC^2$),

$$BCLC^2 = LC^2 - \sigma E, \tag{12}$$

where σ is a user defined weight used to blend the biomechanical model measure with the LC^2 intensity based measure.

4 Results

Registration accuracy was validated on a patient-mimicking phantom of the lumbar spine. Vertebrae L1 to L5 were segmented from patient CT data using ITK-SNAP. The segmented data was converted to a surface model and the spine was printed using a Cimetrix 3D shape printer (Cimetrix Solutions, Oshawa, ON,

Canada). In this model, the natural curvature of the spine is preserved between the patient CT and the printed phantom. The phantom was filled with an agar-gelatine recipe [10], designed to simulate the appearance of soft tissue in US. A high-resolution CT volume (0.46 mm x 0.46 mm x 0.625 mm) and an US volume were acquired. The US volume was reconstructed from a freehand sweep using an L14-5/38 linear-array transducer (Ultrasonix, Richmond, BC, Canada) operating at 6.6 MHz with a depth of 5.5 cm. The probe was tracked using an Optotrack Certus System (Northern Digital Inc., Waterloo, ON, Canada) and calibrated using an N-wire US phantom[11]. Registration was performed on a Dell Precision 690, 2×2.33 GHz Intel Xeon Quad-core CPU and 16GB of RAM.

The phantom CT and US data were aligned to the gold standard, determined by fiducial markers placed on the exterior of the phantom box. The registration was performed on vertebrae L2-L4 and the US volume was cropped to correspond, as seen in Figure 3. The middle vertebrae were chosen for registration as they contained overlap from other vertebrae at the facet joints.

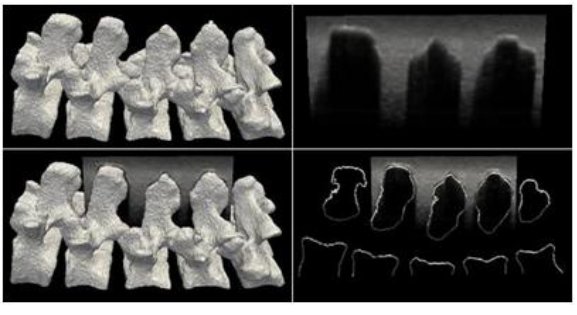


Fig. 3. Surface model of L1 to L5 (top left); US slice of L2 to L4 (top right). Overlaid US with surface model (bottom left) and corresponding extracted slice (bottom right).

One hundred registrations of the CT and US data were performed on the phantom with initial misalignment ranging from 0-20mm Target Registration Error (TRE). The CT volume was misaligned by a random transform chosen from a uniform distribution of 10 mm translation along each axis and 10 rotation about each axis. Each vertebra was further misaligned by individually applying a random transform using a uniform distribution of 5 mm translation along each axis and 5 rotation about each axis. We chose this distribution to ensure that the registration capture range is greater than that of a clinical setting. The registrations were repeated for σ values of 0, 0.5, 1 and 2.

Registration accuracy was determined by its ability to recover to the gold standard alignment, and is reported as the mean TRE calculated from the misalignment of the eight corners of the volume bounding box. Registration results are presented in Table 1 and an example of the initial misalignment and the final registration is displayed in Figure 4. TRE is calculated for each vertebra and the overall error is calculated as the mean error across the entire volume. A registration is considered failed if the final TRE is greater than 3mm.

Table 1. Final TRE for vertebrae L2-L4 and the mean error of the volume are presented for all successful registrations. Success rate (SR) is defined as the percentage of registrations where the overall final TRE is less than 3mm. SR is presented for all registration and for registrations with initial TRE of less than 10mm.

σ	L2 (mm/std)	L3 (mm/std)	L4 (mm/std)	Overall (mm/std)	SR (%)	SR: ITRE < 10mm (%)
0.0	1.49/0.57	1.83/1.02	1.93/0.85	1.75/0.46	61.0	77.8
0.5	1.82/0.36	1.62/0.54	2.47/0.60	1.97/0.30	82.0	98.6
1.0	2.19/0.34	1.79/0.46	3.04/0.44	2.34/0.21	80.0	94.4
2.0	2.46/0.21	1.58/0.28	4.12/0.38	2.71/0.13	72.0	84.8

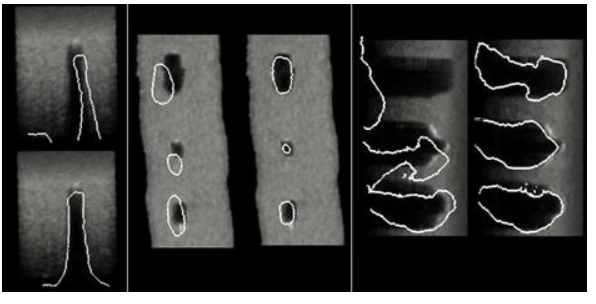


Fig. 4. Transverse (Left), Coronal (Center) and Sagittal (Right) slices of the original US volume overlaid with the bone contours in the misaligned CT volumes and in the registered CT volumes. The transverse slice is taken from the center vertebra L2 (Top).

5 Discussion and Conclusion

In this work we presented an US to CT registration technique for the lumbar spine that successfully registered 98.6% of volumes with initial misalignments of up to 10mm. The registration algorithm applied iterative US simulation from CT images while allowing independent motion of each vertebra. A biomechanical model was introduced to represent the intervertebral link and the system energy was calculated based on the relative transforms between the vertebrae. Integration of a biomechanical model to constrain the registration greatly improved the consistency of the algorithm. When the biomechanical model was combined with the LC^2 metric with a σ weight of 0.5, the algorithm produced the best results: an overall accuracy of 1.97 mm, a failure rate across all registrations of 18% and a failure rate, for registration with initial misalignment less than 10 mm, of 1.4%. The registration technique was tested on patient mimicking phantom that were faithful in representing a realistic curvature of the spine (L1-L5). One hundred tests were performed where each vertebrae was misaligned between 0 mm and 20 mm. The tests were repeated for various weights for integrating the biomechanical model. Using a σ value of 0.5, 82% of all registrations were successful, while 98.6% of tests with initial misalignment of less than 10mm were successfully registered. Increasing the weight of the biomechanical model, to σ equal to

1 or higher, increased the failure rate of the registration. Similarly we observe that while the inclusion of a biomechanical model improves the success rate of a registration, it can also decrease the accuracy of those successful registrations. This can be explained by the fact that the model used is an approximation and not specific to the given patient. The TRE values presented are calculated based on the misalignment of the bounding box corners for each sub-volume. We believe this is a worst case estimate of the error and the error found at the facet joints (approximately centre of the box) will be lower.

The mean run time of our registrations was 57.4min. To reduce this to a more clinically acceptable run time, we have begun the implementation of this algorithm in GPU. Our preliminary implementation, running in CUDA 1.1 on an Nvidia GTX285 graphics processor, reduced the registration runtime to 537s, a 6.4x improvement. In our future work we plan to optimize the GPU implementation and extend the registration to the full lumbar spine, L1-L5. In addition, we also plan to test the registration on real patient data.

References

1. Winter, S., Brendel, B., Pechlivanis, I., Schmieder, K., Igel, C.: Registration of ct and intraoperative 3-d ultrasound images of the spine using evolutionary and gradient-based methods. Evol. Comput. 12(3), 284–296 (2008)
2. Penney, G., Barratt, D., Chan, C., Slomczykowski, M., Carter, T., Edwards, P., Hawkes, D.: Cadaver validation of intensity-based ultrasound to ct registration. Medical Image Analysis 10(3), 385–395 (2006)
3. Wein, W., Khamene, A., Clevert, D.-A., Kutter, O., Navab, N.: Simulation and fully automatic multimodal registration of medical ultrasound. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 136–143. Springer, Heidelberg (2007)
4. Wein, W., Brunke, S., Khamene, A., Callstrom, M.R., Navab, N.: Automatic ct-ultrasound registration for diagnostic imaging and image-guided intervention. Medical Image Analysis 12(5), 577–585 (2008)
5. Shams, R., Hartley, R.I., Navab, N.: Real-time simulation of medical ultrasound from CT images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 734–741. Springer, Heidelberg (2008)
6. Reichl, T., Passenger, J., Acosta, O., Salvado, O.: Ultrasound goes gpu: real-time simulation using cuda, vol. 7261, p. 726116. SPIE (2009)
7. Kutter, O., Karamalis, A., Wein, W., Navab, N.: A gpu-based framework for simulation of medical ultrasound, vol. 7261, p. 726117. SPIE (2009)
8. Gill, S., Mousavi, P., Fichtinger, G., Pichora, D., Abolmaesumi, P.: Group-wise registration of ultrasound to ct images of human vertebrae, vol. 7261, p. 72611O. SPIE (2009)
9. Desroches, G., Aubin, C.E., Sucato, D.J., Rivard, C.H.: Simulation of an anterior spine instrumentation in adolescent idiopathic scoliosis using a flexible multi-body model. Med. Bio. Eng. Comput. 45(8), 759–768 (2007)
10. Madsen, E.L., Hobson, M.A., Shi, H., Varghese, T., Frank, G.R.: Tissue-mimicking agar/gelatin materials for use in heterogeneous elastography phantoms. Phys. Med. Biol. 50(23), 5597–5618 (2005)
11. Chen, T.K., Thurston, A.D., Moghari, M.H., Ellis, R.E., Abolmaesumi, P.: A real-time ultrasound calibration system with automatic accuracy control and incorporation of ultrasound section thickness, vol. 6918, p. 69182A. SPIE (2008)

A General PDE-Framework for Registration of Contrast Enhanced Images

Mehran Ebrahimi and Anne L. Martel

Department of Medical Biophysics, University of Toronto
Imaging Research, Sunnybrook Health Sciences Centre
Toronto, Ontario, Canada
mehran.ebrahimi@sri.utoronto.ca, anne.martel@sri.utoronto.ca

Abstract. This paper presents a general PDE-framework for registration of contrast enhanced images. The approach directly applies the idea of separating the contrast enhancement term from the images in the regularization terms. In our formulation, we stay consistent with existing non-parametric image registration techniques, however, we carry an additional contrast enhancement term throughout. A mathematically rigorous approach is pursued which can exploit various forms of regularization. In this paper, our experiments are built based on diffusion regularization for both contrast enhancement and the deformation field.

1 Introduction

This paper provides a general mathematical framework to complement a series of algorithms [1,2,3,4] that support registration of contrast enhanced images.

Following Horn-Schunck's seminal work [5] on estimating the motion using optical flow equations, a number of approaches were proposed in the literature [6,3,4] which allowed varying illumination among the two images being registered. The approaches developed in [3,4] were very similar to Horn-Schunck's, except that they separated the illumination change as a linear intensity shift term. More recently, a similar idea has been applied to the registration of contrast-enhanced medical images [2,1]. The common intuition behind all these methods is separating and regularizing the contrast enhancement term from the images in the regularization expression coupled with the motion. Numerically, these approaches propose constructing a very large system of equations directly based on the slightly modified optical flow equations. Iterative solvers (e.g., conjugate gradient method) are used to minimize the objective functional and estimate the deformation vectors. All of these methods used the diffusion penalty [7] function for the deformation field in practice and application of other regularization expressions is not readily available. In this paper, we present a general PDE-framework for registration of contrast enhanced images. A goal of our work is to explain, complete, and extend the described set of methods.

To proceed, we need to rigorously introduce the required background material which will be employed in our formulation. For consistency, we adapt our notations from [7]. In Section 2, we will introduce the mathematical formulation and

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 811–819, 2009.

problem set-up. In Section 3, we derive a PDE with a steady-state solution that corresponds to the solution of the described problem. The discretization and derivation of a numerical scheme for the PDE is followed in Section 4. Finally, we will present various computational experiments and concluding remarks in Sections 5 and 6.

2 Mathematical Formulation

Consider the registration problem of a template image T to a reference image R, where T is a realization of R deformed via a vector field u and the contrast of a portion of this realization is changed via an extra additive term (image) w.

In this paper, the d-dimensional reference and template images are represented by mappings $R, T : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ of compact support. The goal is to find a displacement vector field $u = (u_1, \ldots, u_d)$ and a compactly supported contrast enhancement image $w : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ such that $T_u - w$ is similar to R, in which $T_u = T(x - u(x))$ is the deformed image and $T_u - w = T(x - u(x)) - w(x)$ is the corrected-contrast deformed image. We formulate the registration of a contrasted-enhanced deformed image T to a reference image R as following.

Problem 1. *Given two images R and T, find a deformation u and a contrast enhancement image w, that minimizes*

$$\mathcal{J}[u, w] := \mathcal{D}[R, T; u, w] + \mathcal{H}[u, w]$$

in which $\mathcal{D}$ measures the similarity of $T_u - w$ and R, and $\mathcal{H}$ is a regularization expression on $[u, w]$. Here, assuming positive regularizing parameters $\alpha, \beta \in \mathbb{R}^+$ we express

$$\mathcal{H}[u, w] := \alpha\mathcal{P}[u] + \beta\mathcal{Q}[w]. \tag{1}$$

Also, we use the sum of squares of intensity differences for the similarity measure

$$\mathcal{D}[R, T; u, w] := \frac{1}{2} \left|\left|T_u - w - R\right|\right|^2_{L_2(\Omega)}. \tag{2}$$

Hence, the objective is to minimize

$$\mathcal{J}[u, w] := \frac{1}{2} \left|\left|T_u - R - w\right|\right|^2_{L_2(\Omega)} + \alpha\mathcal{P}[u] + \beta\mathcal{Q}[w].$$

We shall present a mathematical formulation to solve Problem 1. Briefly speaking, we seek necessary conditions for optimality of $[u, w]$ by finding the Gâteaux derivatives of the components of $\mathcal{J}$ with respect to $[u, w]$. This shall provide us with the corresponding Euler-Lagrange equations that will be used to form a PDE which will be solved numerically.

Theorem 1. *Let $d \in \mathbb{N}$, and R, T, and w are d-dimensional real-valued images, i.e., functions from $\mathbb{R}^d \to \mathbb{R}$, $T \in C^2(\mathbb{R}^d)$, $u : \mathbb{R}^d \to \mathbb{R}^d$, $v : \mathbb{R}^d \to \mathbb{R}^{d+1}$, $\Omega :=]0, n[^d$. The Gâteaux derivative of $\mathcal{D}[R, T; u, w]$ is given by*

$$d\mathcal{D}[R, T; u, w; v] = -\int_\Omega \langle \Phi(x, u(x), w(x)), v(x) \rangle_{\mathbb{R}^{d+1}} \, dx,$$

in which $\Phi : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R} \to \mathbb{R}^{d+1}$,

$$\Phi(x, u(x), w(x)) = [T_u(x) - R(x) - w(x)]\big(\nabla T_u(x), 1\big).$$

[see the proof in Appendix 1., Cf. [7] pp. 80.]

Here, we focus on the special case where $\mathcal{P}, \mathcal{Q}$ are diffusion regularization expressions [7,8,9,5,4,3,10,11].

Theorem 2. *Assume $\mathcal{P}$ and $\mathcal{Q}$ are diffusion regularization expressions and the functionals $\mathcal{P}^e$ and $\mathcal{Q}^e$ are respectively extensions of $\mathcal{P}$ and $\mathcal{Q}$, i.e.,*

$$\mathcal{P}^e[(u, w)] := \mathcal{P}[u] := \frac{1}{2}\sum_{j=1}^{d}\int_\Omega \langle \nabla u_j, \nabla u_j\rangle \, dx, \tag{3}$$

$$\mathcal{Q}^e[(u, w)] := \mathcal{Q}[w] := \frac{1}{2}\int_\Omega \langle \nabla w, \nabla w\rangle \, dx. \tag{4}$$

Also, assume that Neumann boundary conditions are imposed, i.e.,

$$\langle \nabla w(x), \overrightarrow{n}(x)\rangle_{\mathbb{R}^d} = \langle \nabla u_j(x), \overrightarrow{n}(x)\rangle_{\mathbb{R}^d} = 0 \quad \textit{for} \quad x \in \partial\Omega \ \textit{and} \quad j = 1, \ldots, d,$$

in which $\overrightarrow{n}$ denotes the outer normal unit vector of $\partial\Omega$ (boundary of Ω). The Gâteaux derivative of $\mathcal{P}^e[(u, w); v]$ and $\mathcal{Q}^e[(u, w); v]$ are respectively

$$d\mathcal{P}^e[(u, w); v] = -\int_\Omega \langle \mathcal{A}[u](x), v(x)\rangle_{\mathbb{R}^{d+1}} \, dx,$$

$$d\mathcal{Q}^e[(u, w); v] = -\int_\Omega \langle \mathcal{B}[w](x), v(x)\rangle_{\mathbb{R}^{d+1}} \, dx$$

where, $$\mathcal{A}[u](x) = (\Delta u_1(x), \ldots, \Delta u_d(x), 0) = (\Delta u(x), 0),$$

$$\mathcal{B}[w](x) = (\underbrace{0, \ldots, 0}_{d \ times}, \Delta w(x)) = (0_{\mathbb{R}^d}, \Delta w(x)).$$

Proof. The result yields applying the Green's formula similar to [7] pp. 138.

Theorem 3. *The Euler-Lagrange equations corresponding to $\mathcal{J} = \mathcal{D} + \alpha\mathcal{P} + \beta\mathcal{Q}$, where $\mathcal{D}$ is defined by Equation (2) and $\mathcal{P}$, $\mathcal{Q}$ are defined by Equations (3,4) respectively are*

$$\Phi(x, u(x), w(x)) + \alpha\mathcal{A}[u](x) + \beta\mathcal{B}[w](x) = 0, \qquad x \in \Omega, \tag{5}$$

with Neumann boundary conditions. These can also be written as

$$[T_u(x) - R(x) - w(x)]\nabla T_u(x) + \alpha\Delta u(x) = 0_{\mathbb{R}^d} \qquad x \in \Omega,$$

$$[T_u(x) - R(x) - w(x)] + \beta\Delta w(x) = 0 \qquad x \in \Omega,$$

$$\langle \nabla w(x), \overrightarrow{n}(x)\rangle_{\mathbb{R}^d} = \langle \nabla u_l(x), \overrightarrow{n}(x)\rangle_{\mathbb{R}^d} = 0, \ l = 1, \ldots, d, \ x \in \partial\Omega.$$

Proof. The result yields using substitution (Cf. [7] pp. 138.) .

3 A Corresponding PDE

There exist various ways to solve Equation (5). A possibility that we pursue here is to formulate the solution as the steady-state solution of a corresponding PDE similar to [7]. We propose

$$\partial_t(u(x,t), s\, w(x,t)) = \Phi(x, u(x,t), w(x,t)) + \alpha\mathcal{A}[u](x) + \beta\mathcal{B}[w](x) \quad x \in \Omega, \;\; t \geq 0,$$

where s is a scale factor. Assuming $\Phi = (f, g)$ the PDE can be written as

$$\partial_t u(x,t) = f(x, u(x,t), w(x,t)) + \alpha\Delta u(x,t), \quad x \in \Omega, \;\; t \geq 0, \tag{6}$$

$$s\;\partial_t w(x,t) = g(x, u(x,t), w(x,t)) + \beta\Delta w(x,t), \quad x \in \Omega, \;\; t \geq 0, \tag{7}$$

$$f(x,u,w) := [T_u(x) - R(x) - w(x)]\nabla T_u(x),$$

$$g(x,u,w) := [T_u(x) - R(x) - w(x)].$$

4 Discretization and Numerical Scheme

To numerically solve the derived PDE in Equations (6,7), we evaluate expressions at discrete time variable $\{t_{k+1}\}$

$$\partial_t u(x,t_{k+1}) = f(x, u(x,t_k), w(x,t_k)) + \alpha\Delta u(x,t_{k+1}), \quad x \in \Omega, \tag{8}$$

$$s\;\partial_t w(x,t_{k+1}) = g(x, u(x,t_{k+1}), w(x,t_{k+1})) + \beta\Delta w(x,t_{k+1}), \quad x \in \Omega. \tag{9}$$

Notice that due to the nonlinearity of f with respect to u, f is evaluated at t_k instead of t_{k+1} in Equation (8) [cf. [7] pp. 80] which translates to applying a fixed-point iteration scheme. However, g is linear with respect to w and t_{k+1} is used consistently in Equation (9). Using a spatial discretization X of Ω that includes n^d voxels (pixels) corresponding to a unit space step in every dimension due to the definition of $\Omega :=]0, n[^d$, and a time step of τ_1, we define for $j = 1, \ldots, d$, and $k \in \mathbb{Z}^*$ (the set of non-negative integers)

$$U_j^k(X) := u_j(X, \tau_1 k) := Discretized(u_j(x, t_k)),$$

$$W^k(X) := w(X, \tau_1 k) := Discretized(w(x, t_k)).$$

Furthermore, $\quad AU_j^k := \Delta u_j(X, \tau_1 k) := Discretized(\Delta u_j(x, t_k)),$

$$AW^k := \Delta w(X, \tau_1 k) := Discretized(\Delta w(x, t_k)),$$

in which $A \in \mathbb{R}^{n^d \times n^d}$ is defined such that

$$AU_j^k \approx \sum_{l=1}^{d} \partial_{x_l,x_l} u_j(X,\tau_1 k) \quad \text{and} \qquad AW^k \approx \sum_{l=1}^{d} \partial_{x_l,x_l} w(X,\tau_1 k).$$

[See Appendix 2 for the precise definition of A.]

Substituting the discretization in the PDEs of Equations (8,9) leads that for $j = 1,\ldots,d,\ k \in \mathbb{Z}^*$

$$\frac{U_j^{k+1} - U_j^k}{\tau_1} = \Big(T(X - U^k(X)) - R(X) - W^k(X)\Big)\partial_j\, T(X - U^k(X)) + \alpha A U_j^{k+1},$$

$$s\frac{W^{k+1} - W^k}{\tau_1} = \Big(T(X - U^{k+1}(X)) - R(X) - W^{k+1}(X)\Big) + \beta A W^{k+1}.$$

Defining $\tau_2 := \tau_1/s$ gives

$$\Big(I - \tau_1\alpha A\Big)U_j^{k+1} = U_j^k + \tau_1\Big(T(X - U^k(X)) - R(X) - W^k(X)\Big)\partial_j\, T(X - U^k(X)),$$

$$\Big((1+\tau_2)I - \tau_2\beta A\Big)W^{k+1} = W^k + \tau_2\Big(T(X - U^{k+1}(X)) - R(X)\Big),$$

where $I \in \mathbb{R}^{n^d \times n^d}$ is the identity matrix. This yields

$$U_j^{k+1} = (I - \tau_1\alpha A)^{-1}[U_j^k + \tau_1(T(X - U^k(X)) - R(X) - W^k(X))\partial_j\, T(X - U^k(X))]$$

$$W^{k+1} = ((1+\tau_2)I - \tau_2\beta A)^{-1}[W^k + \tau_2(T(X - U^{k+1}(X)) - R(X))]. \tag{10}$$

Finally, we use the initialization vectors $W^0 = U_j^0 = 0_{\mathbb{R}^{n^d}}, \quad j = 1,\ldots,d.$

5 Computational Experiments

To better understand the introduced scheme, we start from two 2-dimensional 8-bit, 256×256 images shown in Fig. 1(a1-a2) i.e., $d = 2$, $n = 256$. The intensities of images have also been mapped to [0,1]. The reference image is generated by deforming the template using a finite element model (FEM), where the magnitude of the deformation is shown in Fig. 1b(1), followed by subtracting a contrast enhancement term shown in Fig. 1b(3) from the deformed template image. The contrast enhancement term is generated as a combination of physiological enhancement term taken from a patient study and an artificial gradient term.

We apply the scheme derived in Equation (10) to evaluate the deformation and the contrast enhancement terms given only the reference and template image. In all of these experiments, we fix the number of iterations to 50, choose $\alpha = 0.1$, $\tau_1 = 10^3$, $\tau_2 = 10^{10}$ and vary the parameter β. The result are shown in rows (c-f) of Figure 1. The first four columns in each row respectively corresponds to the

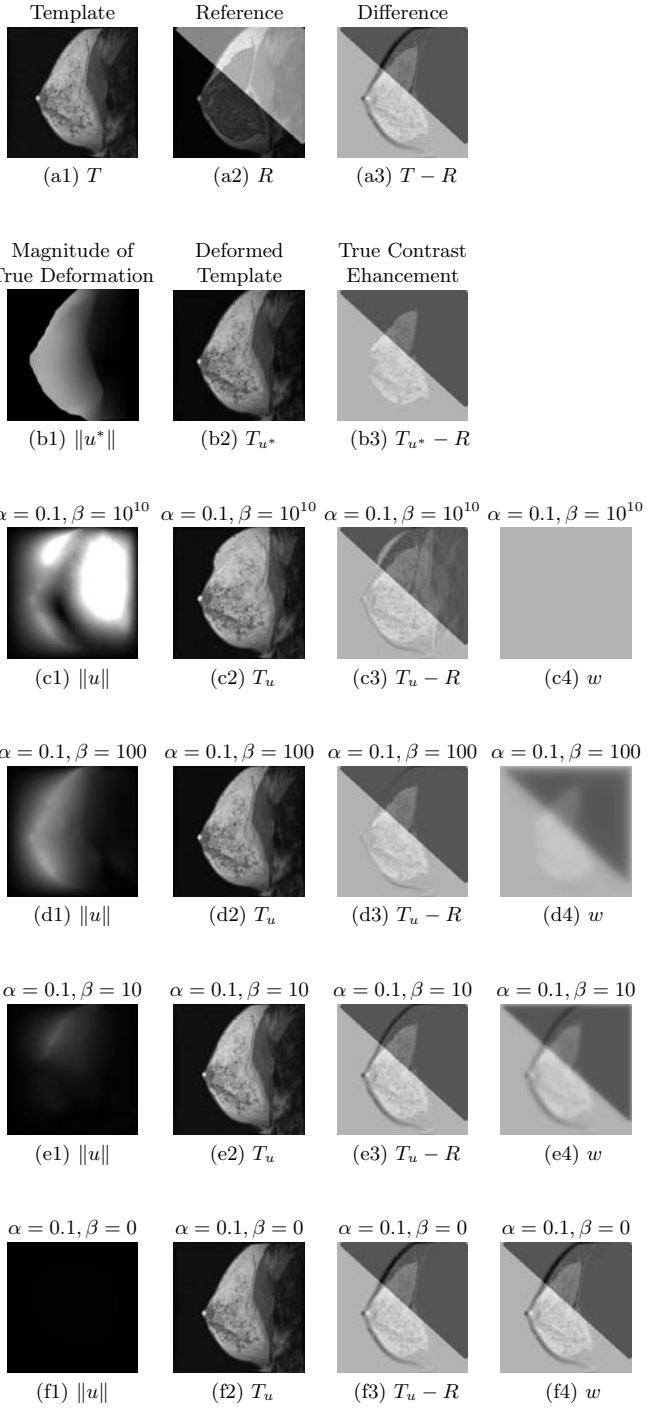


Fig. 1. Computational experiments on simulated data

magnitude of the corresponding deformation shown in a [0-15] gray-level map, the deformed template, the difference between the deformed template and the reference, and the computed contrast enhancement. The employed parameter β for each row is indicated above the Figures.

The case $\beta = 10^{10}$ shown in row (c) corresponds to a diffusion-based registration ignoring the contrast term, while the case $\beta = 0$ in row (f) corresponds to the case that contrast term is not penalized and can vary arbitrarily. It can be observed that among the selected parameters β, the choice of $\beta = 100$ shown in row (d) gives the most reasonable estimation of the motion and the contrast enhancement term.

6 Concluding Remarks

We offered a mathematical framework for registration of contrast enhanced images. Our approach directly applies the idea of separating the contrast enhancement term from the images in regularization [3,4,1,2]. We applied diffusion regularizations for both u and w, where the idea of Adaptive Operator Splitting (AOS) can be simply used to split the operator A of Equation (10) to present a scheme of linear complexity similar to [7]. Finally, the choice of regularizer $\mathcal{H}[u, w]$ in Equation (1) may be replaced with other suitable expressions.

Acknowledgments. This research was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC) in the form of a Postdoctoral Fellowship for Mehran Ebrahimi. This work was also supported by the Terry Fox Foundation for Cancer Research. The authors thank Dr. Ellen Warner (Sunnybrook Health Sciences Centre) and Dr. Kristy Brock (Princess Margaret Hospital) for providing the MR data and FEM simulations.

References

1. Martel, A.L., Plewes, D.B., Froh, M.S., Brock, K.K., Barber, D.C.: Evaluating an optical-flow-based registration algorithm for contrast-enhanced magnetic resonance imaging of the breast. Physics in Medicine and Biology 52(13), 3803–3816 (2007)
2. Barber, D.C., Hose, D.R.: Automatic segmentation of medical images using image registration: diagnostic and simulation applications. Journal of medical engineering and technology 29(2), 53–63 (2005)
3. Gennert, M.A., Negahdaripour, S.: Relaxing the brightness constancy assumption in computing optical flow. 62 Computers, Control and Information Theory AIMEMO975; ADA1874379, MIT (June 1987)
4. Negahdaripour, S., Yu, C.H.: A generalized brightness change model for computing optical flow. In: 1993 IEEE 4th International Conference on Computer Vision, pp. 2–11. IEEE Computer Society Press, Los Alamitos (1993)
5. Horn, B.K.P., Schunck, B.G.: Determining optical flow. Artificial Intelligence 17(1-3; special volume), 185–203 (1981)

6. Weickert, J., Bruhn, A., Brox, T., Papenberg, N.: A survey of variational optic flow methods for small displacement. In: Mathematical Models for Registration and Applications to Medial Imaging, Springer, Heidelberg (2005)
7. Modersitzki, J.: Numerical methods for image registration. Oxford University Press, Oxford (2004)
8. Fischer, B., Modersitzki, J.: Ill-posed medicine-an introduction to image registration. Inverse Problems 24(3), 16 (2008); 034008
9. Fischer, B., Modersitzki, J.: Fast image registration: a variational approach. In: Applied Numerical Analysis and Computational Mathematics, 2004 (NACoM-2003), Cambridge, UK, Germany, vol. 1, pp. A69–A74. Wiley-VCH, Chichester (2004)
10. Thirion, J.P.: Fast non-rigid matching of 3d medical images. In: Medical Robotics and Computer Aided Surgery, Baltimore, pp. 47–54 (1995)
11. Thirion, J.P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Medical Image Analysis 2(3), 243–260 (1998)
12. Ebrahimi, M., Martel, A.L.: Image registration under varying illumination: Hyperdemons algorithm. (preprint, 2009)

Appendix 1: Proof of Theorem 1

Proof. Define $p(x) := (v_1(x), \ldots, v_d(x))$ and $q(x) := v_{d+1}(x)$, i.e., split $v(x)$ to $v(x) = (p(x), q(x))$. Using the Taylor expansion of $T_{u+hp}(x)$ with respect to h at the point $x - u(x)$,

$$T_{u+hp}(x) = T(x - u(x) - hp(x))) = T_u(x) - h\langle \nabla T_u(x), p(x)\rangle_{\mathbb{R}^d} + \mathcal{O}(h^2).$$

Hence, (Cf. [7] pp. 81.)

$$\begin{aligned}
d\,\mathcal{D}\,[R,T;u,w;v] &= \lim_{h\to 0}\frac{1}{h}(\mathcal{D}[R,T;(u,w)+hv] - \mathcal{D}[R,T;u,w]) \\
&= \lim_{h\to 0}\frac{1}{h}(\mathcal{D}[R,T;(u+hp,w+hq)] - \mathcal{D}[R,T;u,w]) \\
&= \lim_{h\to 0}\frac{1}{2h}\int_\Omega (T_{u+hp}(x) - R(x) - (w(x)+hq(x)))^2 \\
&- (T_u(x) - R(x) - w(x))^2\; dx \\
&= \lim_{h\to 0}\frac{1}{2h}\int_\Omega (T_u(x) - h\langle \nabla T_u(x), p(x)\rangle_{\mathbb{R}^d} + \mathcal{O}(h^2) - R(x) - w(x) - hq(x))^2 \\
&- (T_u(x) - R(x) - w(x))^2\; dx \\
&= \lim_{h\to 0}\frac{1}{2h}\int_\Omega 2[T_u(x) - R(x) - w(x)](-h\langle \nabla T_u(x), p(x)\rangle_{\mathbb{R}^d} - hq(x) + \mathcal{O}(h^2))\; dx \\
&= \int_\Omega -[T_u(x) - R(x) - w(x)][\langle \nabla T_u(x), p(x)\rangle_{\mathbb{R}^d} + \langle 1, q(x)\rangle_{\mathbb{R}}]\; dx \\
&= \int_\Omega \langle -[T_u(x) - R(x) - w(x)]\big(\nabla T_u(x), 1\big), v(x)\rangle_{\mathbb{R}^{d+1}}\; dx.
\end{aligned}$$

Appendix 2: Definition of A

$A \in \mathbb{R}^{n^d \times n^d}$ is defined as $A := \sum_{l=1}^{d} A_l$ where $A_l = \underbrace{I \otimes \cdots \otimes I}_{l-1 \text{ times}} \otimes B \otimes \underbrace{I \otimes \cdots \otimes I}_{d-l \text{ times}}$, in which $I \in \mathbb{R}^{n \times n}$ is identity matrix and $\otimes$ denotes the Kronecker product of matrices. The l^{th} factor $B \in \mathbb{R}^{n \times n}$ is an approximation of the second order derivative in only one spatial direction. More precisely, it can be defined as the tridiagonal matrix

$$B = \begin{pmatrix} -2 & 1 & 0 & \dots & 0 \\ 1 & -2 & 1 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 1 & -2 & 1 \\ 0 & \dots & \dots & 1 & -2 \end{pmatrix}.$$

Statistically Deformable 2D/3D Registration for Accurate Determination of Post-operative Cup Orientation from Single Standard X-ray Radiograph

Guoyan Zheng

ISTB, University of Bern, Stauffacherstrasse 78, CH-3014 Bern, Switzerland
guoyan.zheng@ieee.org

Abstract. The widely used procedure of evaluation of cup orientation following total hip arthroplasty using single standard anteroposterior (AP) radiograph is known inaccurate, largely due to the wide variability in individual pelvic orientation relative to X-ray plate. 2D/3D rigid image registration methods have been introduced for an accurate determination of the post-operative cup alignment with respect to an anatomical reference extracted from the CT data. Although encouraging results have been reported, their extensive usage in clinical routine is still limited. This may be explained by their requirement of a CAD model of the prosthesis, which is often difficult to be organized from the manufacturer due to the proprietary issue, and by their requirement of a pre-operative CT scan, which is not available for most retrospective studies. To address these issues, we developed and validated a statistically deformable 2D/3D registration approach for accurate determination of post-operative cup orientation. No CAD model and pre-operative CT data is required any more. Quantitative and qualitative results evaluated on cadaveric and clinical datasets are given, which indicate the validity of the approach.

1 Introduction

Two-dimensional (2D) anteroposterior (AP) pelvic radiographs, despite their inferior accuracy in comparison to three-dimensional (3D) techniques based on computed tomography (CT) [1], are the standard imaging method for the evaluation of cup orientation following total hip arthroplasty (THA) [2], largely due to the simplicity, availability, and minimal expense associated with acquiring these images. While plain pelvic radiographs are easily obtained, their accurate interpretation is complicated by the wide variability in individual pelvic orientation relative to the X-ray plate [1]. In THA, increased pelvic tilt results in significant decreases in apparent prosthetic cup anteversion and vice versa [3].

2D/3D rigid registration methods [4] [5] [6] have been introduced to estimate the post-operative cup alignment with respect to an anatomical reference, which is a plane called the anterior pelvic plane (APP) defined by the anterior superior iliac spines (ASIS) and the pubic tubercles [8]. The common procedure within

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 820–827, 2009.

these methods is to use 2D/3D registration to align both the prosthesis and the pre-operative CT volume to the X-ray image(s), which then allows the position of the implant to be calculated with respect to the anatomical reference plane. Although encouraging results have been reported, their extensive usage in clinical routine is still limited. This may be explained by their requirement of a Computer Aided Design (CAD) model of the prosthesis [4] [5] [6], which often is difficult to be organized from the manufacturer due to the proprietary issue, and by their requirement of a pre-operative CT scan [4] [5] [6], which is not available for most retrospective studies or earlier cup designs. We addressed the former issue in our previous work [7], where a hybrid 2D/3D registration scheme was implemented to estimate a rigid transformation between a pre-operative CT volume and the post-operative X-ray radiograph. No registration between the prosthesis and the X-ray radiograph is required, thus eliminating the necessity of possessing the CAD models of the prosthesis. In this paper, we would like to address the latter issue through 2D/3D reconstruction.

Reconstructing a 3D bone model from single 2D projection image is a challenging task. Furthermore, in our application, no radiograph-specific calibration is available. The only information that we assume to know about the radiograph is the image scale (mm/pixel) and the distance from the focal point to the imaging plane or to the film. As long as the radiograph is acquired in a standardized way, which is preformed in a clinical routine [2], both parameters can be estimated by performing one-time calibration [9]. When only single image is used, it is well-known that the depth information and the scaling factor are correlated with each other. Thus, precise estimation of both of them from single image is difficult, if it is not impossible. However, considering the context of our application, we hypothesize that even a scaled estimation of the pelvic model should provide enough information to determine the post-operative cup orientation. We thus developed a statistically deformable 2D/3D registration approach to instantiate a patient-specific pelvis surface model from the single standard X-ray radiograph and then used the instantiated surface model to determine the post-operative cup orientation.

This paper is organized as follows. Section 2 briefly presents the construction of the statistical shape model. Section 3 describes the statistically deformable 2D/3D registration approach. Section 4 presents the experimental results, followed by the conclusions in Section 5.

2 Construction of the Statistical Shape Model of the Pelvis

We chose the point distribution model (PDM) as the representation of our statistical shape model of the pelvis. The pelvic PDM used in this paper was constructed from a training database consisted of 14 segmented binary volumes (12 of them were segmented from CT scans of dry bones and the rest 2 were segmented from patient CT scans) where the sacrum was removed from each dataset. Demon's algorithm [10] as implemented in MedINRIA [11] was used to

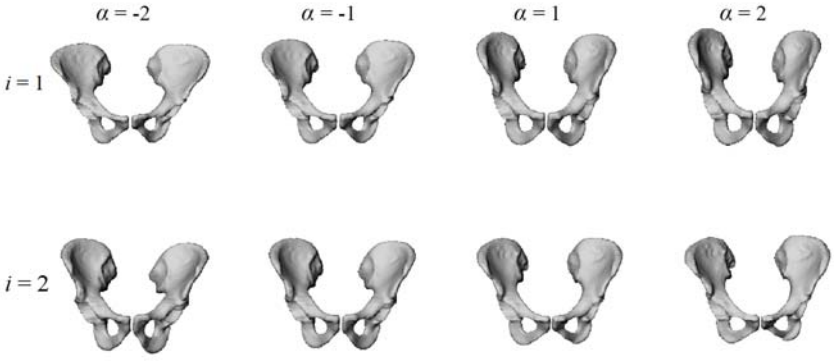


Fig. 1. The first two eigen modes of variation of our PDM of the pelvis. The shape instances were generated by evaluating $\bar{x} + \alpha\sigma_i \mathbf{p}_i$ with $\alpha \in \{-2, -1, 1, 2\}$.

estimate the dense deformation fields between the reference binary volume and the other 13 binary volumes. Each estimated deformation field was then used to displace the positions of the vertices on the reference surface model, which was constructed from the reference binary volume, to the associated target volume. We thus obtained 14 surface models with established correspondences.

Following the alignment, the PDM is constructed as follows. Let $\mathrm{x}_i, i = 0, 1, ..., m-1$, be m (here m=14) members of the aligned training surfaces. Each member is described by a vectors x_i with N (here N=24994) vertices:

$$\mathrm{x}_i = \{x_0, y_0, z_0, x_1, y_1, z_1, ..., x_{N-1}, y_{N-1}, z_{N-1}\} \tag{1}$$

The PDM is obtained by applying principal component analysis.

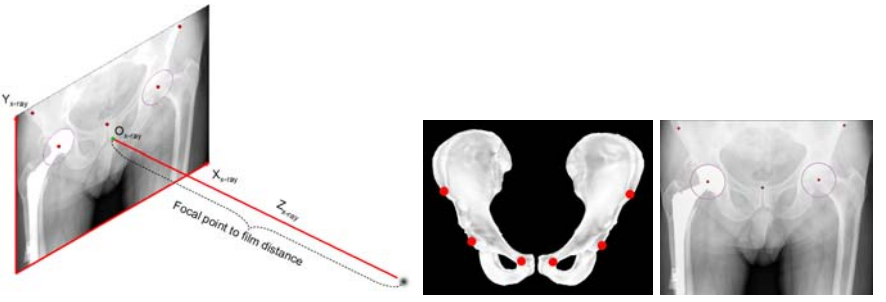
$$\begin{aligned} D &= ((m-1)^{-1}) \cdot \textstyle\sum_{i=0}^{m-1} (\mathrm{x}_i - \bar{\mathrm{x}})(\mathrm{x}_i - \bar{\mathrm{x}})^T \\ P &= (\mathbf{p}_0, \mathbf{p}_1, ...); D \cdot \mathbf{p}_i = \sigma_i^2 \cdot \mathbf{p}_i \end{aligned} \tag{2}$$

where $\bar{\mathrm{x}}$ and D are the mean vector and the covariance matrix, respectively.

Fig 1 shows the variability captured by the first two modes of variation of our PDM.

3 Statistically Deformable 2D/3D Registration

Our single image based surface model reconstruction technique is based on the algorithm that we introduced in [12], which combines statistical instantiation and regularized shape deformation with an iterative image-to-model correspondence establishing algorithm. The image-to-model correspondence is established using a non-rigid 2D point matching process, which iteratively uses a symmetric injective nearest-neighbor mapping operator and 2D thin-plate splines based deformation to find a fraction of best matched 2D point pairs between features extracted from the X-ray images and the projections of the apparent contours



(a) (b) (c)

Fig. 2. (a) The radiograph coordinate system and the cone-beam projection model; (b) Landmarks extracted from the mean model of the PDM; and (c) landmarks extracted from radiograph

extracted from the 3D model. The obtained 2D point pairs are then used to set up a set of 3D point pairs such that we turn a 2D/3D reconstruction problem to a 3D-3D one. The 3D/3D reconstruction problem is then solved optimally in three sequential stages including iterative scaled rigid registration, statistical instantiation, and regularized shape deformation. For details, we refer to our previous works [12].

In our previous work, we asked for 2 or more X-ray images as the input and that all images should be calibrated. However, these requirements are the conditions for the application in our previous work rather than the constraints to our algorithm. Actually, the algorithm that we introduced in [12] can be directly applied to single image, as long as at least four non-colinear point pairs are found. Due to the fact that only single image is available, we may not obtain precise reconstructions as those reported in our previous work, but rather scaled reconstructions. Similar to the situation when multiple images are used, the convergence of the single image based 2D/3D reconstruction also depends on the initialization. Thus, in the following we focus on the establishment of the projection geometry of the input radiograph, and on a landmark-based scaled registration for initializing the single image based 2D/3D reconstruction.

3.1 Establishment of Projection Geometry

The local coordinate reference and the cone-beam projection model of the radiograph is established as follows (see Fig. 2(a) for details). The image center is taken as the coordinate origin. The X-axis and the Y-axis of the image are taken as the X-axis and the Y-axis of the local coordinate reference of the radiograph. The central projection line is perpendicular to the radiograph plane and its opposite direction is regarded as Z-axis.

3.2 Landmark-Based Scaled Registration for Initialization

Initialization here means to estimate the initial scale and the rigid transformation between the mean model of the PDM and the input radiograph. For

this purpose, we have adopted an iterative landmark-to-ray scaled registration. The five anatomical landmarks that we used here are left and right ASIS, left and right acetabular centers, and pubic symphysis. Their positions on the mean model of the PDM are obtained through point picking (for left and right ASIS, and pubic symphysis) or sphere fitting (for left and right acetabular centers), while their positions on the radiograph are defined through interactive picking (for the projections of left and right ASIS, and pubic symphysis) or circle fitting (for the projections of left and right acetabular centers) (see Fig. 2(b) and 2(c) for details).

Let us denote those landmarks defined on the mean model of the PDM, i.e., the left and the right acetabular centers, the pubic symphysis, and the middle points of the left and the right ASIS, as $v^1_{Mean}, v^2_{Mean}, v^3_{Mean}$, and v^4_{Mean}, respectively; and their corresponding landmarks interactively picked from the radiograph as $v^1_{X-ray}, v^2_{X-ray}, v^3_{X-ray}$, and v^4_{X-ray} (v^4_{X-ray} is the middle point of the projections of the left and the right ASIS), respectively. And for each X-ray landmark, we can calculate a projection ray emitting from the focal point to the landmark. We then calculate the length between v^1_{Mean} and v^2_{Mean} and denote it as $l^{1,2}_{Mean}$. Using the known image scale, we also calculate the length $l^{1,2}_{X-ray}$ between v^1_{X-ray} and v^2_{X-ray}. Then, we do:

Data Preparation. In this step, we assume that the line connecting the acetabular centers is parallel to the AP pelvic radiograph plane and is certain distance away from the imaging plane or the film (in all the experiments reported in this paper, we used a fixed distance of 150 mm). Using this assumption and the correspondences between the landmarks defined in the CT volume and those picked from the radiograph, we can first compute two points $\bar{v}^1_{X-ray}$ and $\bar{v}^2_{X-ray}$ on the projection rays of v^1_{X-ray} and v^2_{X-ray}, respectively, which satisfy:

$$\bar{v}^1_{X-ray}\bar{v}^2_{X-ray} // v^1_{X-ray}v^2_{X-ray}; \; and \; |\bar{v}^1_{X-ray} - \bar{v}^2_{X-ray}| = l^{1,2}_{X-ray} \times \frac{F-d}{F} \quad (3)$$

where F is the known distance from the focal point to the imaging plane and d is the assuming distance from the acetabular centers to the imaging plane.

The current scale s between the mean model and the input image is then estimated as,

$$s = |\bar{v}^1_{X-ray} - \bar{v}^2_{X-ray}|/l^{1,2}_{Mean} \quad (4)$$

Using s, we scale all landmark positions on the mean model and denote them as $\{\bar{v}^i_{Mean}$; i=1,2,3, 4$\}$. We then calculate the distances from $\bar{v}^3_{Mean}$ and $\bar{v}^4_{Mean}$ to line $\bar{v}^1_{Mean}\bar{v}^2_{Mean}$ and denote it as $\bar{l}^{3,1-2}_{Mean}$ and $\bar{l}^{4,1-2}_{Mean}$, respectively.

Next we find two points, point $\bar{v}^3_{X-ray}$ on the projection ray of v^3_{X-ray} whose distance to the line $\bar{v}^1_{X-ray}\bar{v}^2_{X-ray}$ is equal to $\bar{l}^{3,1-2}_{Mean}$, and point $\bar{v}^4_{X-ray}$ on the projection ray of v^4_{X-ray} whose distance to the line $\bar{v}^1_{X-ray}\bar{v}^2_{X-ray}$ is equal to $\bar{l}^{4,1-2}_{Mean}$. A paired-point matching [13] based on $\{\bar{v}^i_{Mean}$; i=1,2,3,4$\}$ and $\{\bar{v}^i_{X-ray}$; i=1,2,3,4$\}$ is used to calculate a updated scale s_0 and a rigid transformation $\bar{T}^{X-ray}_{Mean}$ (see Fig. 3(a) for details). From now on, we assume that all information defined in the mean model coordinate frame has been transformed into the

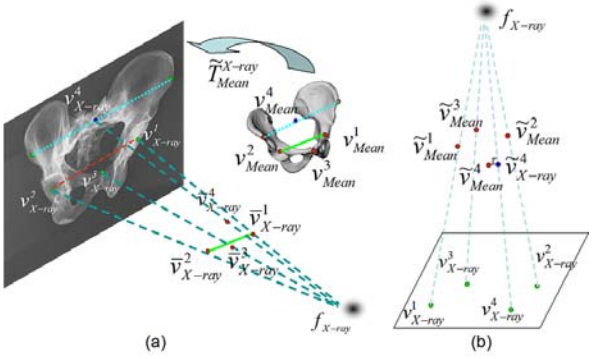


Fig. 3. Iterative landmark-to-ray registration. (a) schematic view of data preparation; and (b) schematic view of finding 3D point pairs.

radiograph coordinate frame using s_0 and $\bar{T}^{X-ray}_{Mean}$. We denote the transformed mean model landmarks as $\{\tilde{v}^i_{Mean}\}$.

Iteration. The following steps are iteratively executed until convergence:

1. For a point $\tilde{v}^i_{Mean}$, we find a point on the corresponding projection ray of v^i_{X-ray} which has the shortest distance to the point $\tilde{v}^i_{Mean}$ and denote it as $\tilde{v}^i_{X-ray}$ (see Fig. 3(b)). We then perform a paired-point matching [13] using the extracted point pairs to compute a scale $\tilde{s}$ and a rigid transformation $\Delta\tilde{T}^{X-ray}_{Mean}$.
2. We update the mean model coordinate frame using $\tilde{s}$ and $\Delta\tilde{T}^{X-ray}_{Mean}$.

3.3 2D/3D Reconstruction

The estimated scale and the rigid transformation between the mean model and the input image are then treated as the starting values for the algorithm that we introduced in [12]. As a feature-based 2D/3D reconstruction approach, our algorithm requires a pre-requisite image feature extraction. In this paper, observing the superimposition of the projections of different bone structures around the pelvis and the post-operative characteristic of the X-ray radiograph, we opt for an interactive way to identify contours of the pelvis. We thus developed a program allowing the user to define up to eight contours by interactively picking points from the radiograph. Each contour is then interpolated by a cubic-spline to have the same resolution as the image resolution. The extracted contours are then used together with the initial estimation of the scale and the rigid transformation as the input to our PDM based 2D/3D reconstruction scheme for an accurate reconstruction of a surface model of the pelvis. Fig. 4 shows different stages of reconstruction of a patient-specific surface model of the pelvis from single standard X-ray radiograph of a patient. The reconstructed surface model of the pelvis can then be used to determine the post-operative cup orientation.

(a) (b) (c) (d)

Fig. 4. (a) the image contours (white line); (b) establishment of the initial image-to-model correspondences (yellow points: projections of the apparent contours extracted from the mean model; green lines: visualization of the correspondences); (c) the result of the iterative scaled registration; and (d) the final reconstructed surface model

Table 1. Difference between the estimated results estimated and the ground truths

angle	cadaver_01	cadaver_02	cadaver_03	cadaver_04	Patient_01	Patient_02	Mean
anteversion (^{o})	1.0	4.4	3.0	1.0	1.5	1.2	2.0 ± 1.4
inclincation (^{o})	0.9	1.1	0.8	0.7	0.8	2.0	1.4 ± 0.5

4 Experiments and Results

We designed and conducted experiments on 4 cadaveric pelvis datasets and 2 patient datasets (one male and one female. Note: none of them has been included for constructing the PDM) to validate the present approach. Each dataset contains a post-operative X-ray radiograph and a post-operative CT volume. The cup orientations extracted from the associated post-operative CT volume are regarded as the ground truths. The results obtained by the present approach are compared to the associated ground truths to estimate the measurement errors. Two X-ray machines were used to acquire the X-ray radiograph. The X-ray radiographs for all 4 cadaveric pelvis were acquired by one X-ray machine with a focal point to film distance of 1200 mm and a pixel size of 0.143 mm while the X-ray radiographs for two patients were acquired by the other X-ray machine with a focal point to film distance of 1016 mm and a pixel size of 0.17 mm.

The results of our validation experiment are presented in Table 1. Compared to the ground truths, differences of $2.0^o \pm 1.4^o$ were found for the anteversion and differences of $1.4^o \pm 0.5^o$ were found for the inclination. Although the anteversion measurement errors are slightly bigger than those based on pre-operative CT scans [7], they are still in the acceptable range according to Kalteis et al. [1].

5 Conclusions

In this paper, we presented a statistically deformable 2D/3D registration approach to instantiate a patient-specific pelvis surface model from single standard X-ray radiograph and then used the reconstructed model to determine the post-operative cup orientation. We designed and conducted feasibility experiments to validate the present approach. Our experimental results demonstrate that it

is feasible to reconstruct a patient-specific model from single standard X-ray radiograph for accurate determination of post-operative cup orientation.

References

1. Kalteis, T., Handel, M., Herold, T., Perlick, L., Paetzel, C., Grifka, J.: Position of the acetabular cup-accuracy of radiographic calculation compared to CT-based measurement. Eur. J. Radiol. 58, 294–300 (2006)
2. Della Valle, C.J., Kaplan, K., Jazrawi, A., Ahmed, S., Jaffe, W.L.: Primary total hip arthroplasty with a flanged, cemented all-polyethylene acetabular component. J. Arthroplasty 19, 23–26 (2004)
3. Sellers, R.G., Lyles, D., Dorr, L.D.: The effect of pelvic rotation on alpha and theta angles in total hip arthroplasty. Contemp. Orthop. 17, 67–70 (1988)
4. Blendea, S., Eckman, K., Jaramaz, B., Levison, T.J., DiGioia, A.M.: Measurements of acetabular cup position and pelvic spatial orientation after total hip arthroplasty using computed tomography/radiography matching. Computer Aided Surgery 10, 37–43 (2005)
5. Jaramaz, B., Eckman, K.: 2D/3D registration for measurement of implant alignment after total hip replacement. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 653–661. Springer, Heidelberg (2006)
6. Penney, G.P., Edwards, P.J., Hipwell, J.H., Slomczykowski, M., Revie, I., Hawkes, D.J.: Postoperative calculation of acetabular cup position using 2D - 3D registration. IEEE T. Biomed. Eng. 54, 1342–1348 (2007)
7. Zheng, G., Steppacher, S., Zhang, X., Tannast, M.: Precise estimation of postoperative cup alignment from single standard X-ray radiograph. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 951–959. Springer, Heidelberg (2007)
8. Jaramaz, B., DiGioia, A.M., Blackwell, M., Nikou, C.: Computer assisted measurement of cup placement in total hip replacement. Clin. Orthop. Rel. Res. 354, 70–81 (1998)
9. The, B.: Digital radiographic peroperative planning and postoperative monitoring of total hip replacements. Doctoral dissertations, University Medical Center Groningen, the Netherlands (2006)
10. Thirion, J.-P.: Image matching as a diffusion process: an analogy with Maxwell's demons. Med. Image Anal. 2, 243–260 (1998)
11. Toussaint, N., Souplet, J.-C., Fillard, P.: MedINRIA: DT-MRI processing and visualization software. In: MICCAI 2007 Workshop on Interaction in Medical Image and Visualization (2007)
12. Zheng, G., Ballester, M.A.G., Styner, M.A., Nolte, L.-P.: Reconstruction of patient-specific 3D bone surface from 2D calibrated fluoroscopic images and point distribution model. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 25–32. Springer, Heidelberg (2006)
13. Veldpaus, F.E., Woltring, H.J., Dortmans, L.J.: A least-square algorithm for the equiform transformation from spatial marker coordinates. J. Biomech. 21, 45–54 (1988)

A Novel Intensity Similarity Metric with Soft Spatial Constraint for a Deformable Image Registration Problem in Radiation Therapy

Ali Khamene[1], Darko Zikic[2], Mamadou Diallo[1], Thomas Boettger[3], and Eike Rietzel[4]

[1] Imaging and Visualization Dept., Siemens Corporate Research, Princeton NJ, USA
[2] Computer Aided Medical Procedures (CAMP), TU Munich, Germany
[3] Oncology Care Systems, Siemens Healthcare Sector, Heidelberg, Germany
[4] Particle Therapy, Siemens Healthcare Sector, Erlangen, Germany

Abstract. In this paper we propose a novel similarity metric and a method for deformable registration of two images for a specific clinical application. The basic assumption in almost all deformable registration approaches is that there exist explicit correspondences between pixels across the two images. This principle is used to design image (dis)similarity metrics, such as sum of squared differences (SSD) or mutual information (MI). This assumption is strongly violated, for instance, within specific regions of images from abdominal or pelvic section of a patient taken at two different time points. Nevertheless, in some clinical applications, it is required to compute a smooth deformation field for all the regions within the image including the boundaries of such regions. In this paper, we propose a deformable registration method, which utilizes a priori intensity distributions of the regions delineated on one of the images to devise a new similarity measure that varies across regions of the image to establish a smooth and robust deformation field. We present validation results of the proposed method in mapping bladder, prostate, and rectum contours of computer tomography (CT) volumes of 10 patients taken for prostate cancer radiotherapy treatment planning and verification.

1 Introduction and Background

We address the problem of registering two images and determining the dense deformation field mapping one image to the other. Registration of pairs of medical images (2D or 3D) has been extensively studied (see [9]). In all these approaches, the main assumption is that an image (dis)similarity metric can be established, which can evaluate the quality of a deformation field. Furthermore, the deformation field is required to be continuous, smooth and invertible, so that every pixel in image one (fixed or target) maps to exactly one pixel in image two (moving or template), and vice-versa. The smoothness property can be enforced through regularization of the dense deformation field assuming diffusive [5], elastic [1], viscous fluid [3], or splines [11] properties.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 828–836, 2009.

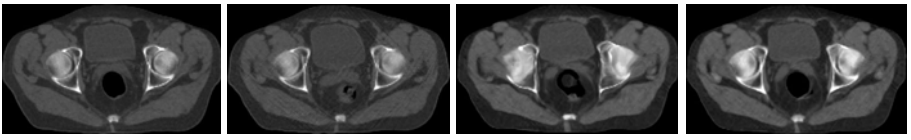


Fig. 1. Figure depicts from left to right, a planning CT slice, a treatment CT slice, the warped version of a treatment CT slice based on a conventional method with SSD as the similarity measure, and the same image done based on the proposed method

A problem arises when, for some parts within the images, no correspondences can be established across the two. This happens quite often while registering medical images. One example is during the registration of CT images from a male pelvis taken for prostate cancer radiotherapy. Rectum and intestine (bowel) content may change drastically from the planning to the treatment sessions. This makes the process of establishing correspondences impossible and more importantly meaningless. There are two problems with the conventional approaches. First is that intensity based similarity metric can not be used for these specific (special) regions within these images. This is mainly due to excessive penalizing effects that they impose on the deformation field within these regions (see Figure 1). If used, this causes erroneous results for the deformation field within these regions. Second, the regularization constraint, and the associated smoothing effect it has, causes the erroneous deformation field to spread to the neighboring areas. The combination intensifies the problem and results an overall unsatisfactory deformation field. This problem is also identified in Foskey et. al. [6], where a two step registration is proposed to deal with the presence of gas in the rectum. We believe that a good approach for solving this problem should be based on a unified variational formulation that can recover the dense deformation field and also deal with these special regions. Joint registration and segmentation approaches have also been proposed, for example in [12] and [4]. However, in [12], the approach only produces deformation fields on the border of the regions (contour in 2D or surface in 3D). In [4] also there is no explicit consideration for the special regions, where no match can exist. We propose a variational registration framework that incorporates statistical intensity constraints on the regions suspected to not have a one-to-one correspondences within the other image. We formulate the problem as a non-linear optimization to solve for the dense deformation, which is based on a combination of a conventional and the region based intensity similarity metric and a classic regularization constraint to enforce smoothness globally on the deformation field.

2 Deformable Registration: Problem Formulation

We consider the problem of deformable registration, where an unknown transformation $\mathcal{T} : \Omega_m \rightarrow \Omega_f$ which maps the intensity images I_m: $\Omega_m \rightarrow \Re$ (i.e., moving) to $I_f : \Omega_f \rightarrow \Re$ (i.e., fixed) defined over the space Ω ($\Omega \in \Re^k$ for $k = 2$ or 3). We assume to have outlines of special regions specified on the fixed image.

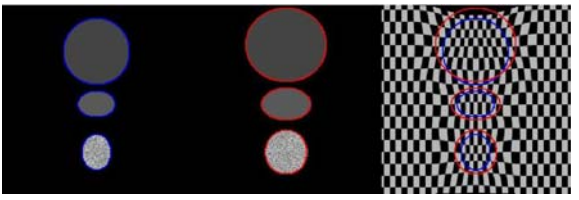


Fig. 2. Figure shows two images (from left) with three deformable objects and their corresponding contours. The lower object in both images contains random distribution of intensities inside, where the pixel-to-pixel correspondences does not exist across the two. Third image (right) shows the desired deformation, which maps the region boundaries and provides a smooth deformation within and on the boundaries of the lower object.

We compartmentalize Ω_f to a set of non-overlapping regions Φ_f^i for $i \in [0, n-1]$, where Φ_f^0 denotes the region, where matching pairs can be found (i.e., majority of image pixels) and Φ_f^i for $i \neq 0$, where matching pairs can not be established. We assume the contours enclosing regions Φ_f^i for $i \neq 0$ are known for I_f and denoted by $C_i^f \in \Omega_f$. We are looking for the transformation $\mathcal{T}$ with the following properties (see Figure 2):

- For any point in the fixed image with in the region Φ_f^0, there should be a corresponding point in the moving image (i.e., $\mathbf{x}_f = \mathcal{T}(\mathbf{x}_m)$ for $\mathbf{x}_f \in \Phi_f^0$).
- A point on the contour of C_i^f for $i \neq 0$ should map to a point on a closed contour denoted as C_i^m for $i \neq 0$ (i.e., $C_i^f = \mathcal{T}(C_i^m)$ for $i \neq 0$). It is important to emphasize that this is only a boundary constraint and does not enforce having correspondences within these regions.
- The transformation $\mathcal{T}$ should be smooth.

3 Proposed Solution

We use the classic two term energy functional for solving this problem. We reformulate the transformation $\mathcal{T}$ as $\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathbf{u}(\mathbf{x})$, where $\mathbf{u}$ is the displacement (deformation) field. We use various image (dis)similarity terms (i.e., $\mathcal{M}_{\Phi_i}$) depending on the region and one overall deformation smoothness constraint (i.e., $\mathcal{R}$), as follows:

$$\hat{\mathbf{u}} = \arg\min_{\mathbf{u}} \sum_i \beta_i \mathcal{M}_{\Phi_i}(I^f, I^m) + \alpha \int_\Omega \mathcal{R}(\mathbf{u}(\mathbf{x}))d\mathbf{x}, \tag{1}$$

where β_i for $i \in [0, n-1]$ needs to be specified empirically. Furthermore, α signifies the level of smoothness in the deformation field. For region Φ_0, where the correspondences can be found, we could use conventional similarity metrics such as MI [9] or SSD as follows:

$$\mathcal{M}_{\Phi_0}(I^f, I^m) = \int_{\Phi_0} |I^f(\mathbf{x}) - I^m(\mathbf{x} + \mathbf{u}(\mathbf{x}))|^2 d\mathbf{x}. \tag{2}$$

For other regions, we propose to use region based similarity metric with no specific spatial information. We assume that the a priori intensity distributions for these regions are known. This information for example can be established based on outlining various regions within several training data sets. Let us assume that the a priori intensity distribution of the region Φ_i for $i \in [1, n-1]$ is denoted by p_i, we can then define a log likelihood based regional energy constraint as follows:

$$\mathcal{M}_{\Phi_i}(I^f, I^m) = -\sum_i \int_{\Phi_i} log(p_i(I^m(\mathbf{x} + \mathbf{u}(\mathbf{x})))d\mathbf{x}. \tag{3}$$

It is noteworthy to mention that there is no notion of correspondences present in this equation. The regional energy is minimized for the deformation field, which maps the certain area within the moving image into the region with a known intensity distribution. For example, in a special case, where p_i is a Gaussian distribution, the equation 3 is favoring deformation fields that cause the intensity mean of the mapped region be close to the mean of the a priori distribution. The intensity distribution could either be estimated based on the I_f alone, or it could be estimated using the same image regions from a number patients' scans. The latter is a more reliable option, however can only be used for quantitative imaging modalities such as CT. Finally, as the global deformation regularization, we simply use the diffusion criterion $\mathcal{R}(\mathbf{u}) = trace(\nabla \mathbf{u}^\top \nabla \mathbf{u})$ where $\top$ denotes the transpose operation. Any other regularizer, such as linear-elastic or curvature [10], could be used as well.

3.1 Numerical Treatment

The optimal deformation field minimizing the energy functional (1) is the solution of the Euler-Lagrange equation of $\alpha \Delta \mathbf{u}(\mathbf{x}) = \mathbf{f}(\mathbf{x})$, with certain boundary conditions [10]. In the proposed case, the right hand side in this equation or the image force is modified according to the regions of the fixed image as follows:

$$\begin{aligned}\mathbf{f}(\mathbf{x}) = \Big(&\beta_0 \mathcal{X}_{\phi_0}(\mathbf{x})(I^f(\mathbf{x}) - I^m(\mathbf{x} + \mathbf{u}(\mathbf{x})))+ \\ &\textstyle\sum_i \beta_i \mathcal{X}_{\phi_i}(\mathbf{x}) \frac{\partial p_i(I^m(\mathbf{x}+\mathbf{u}(\mathbf{x})))}{p_i(I^m(\mathbf{x}+\mathbf{u}(\mathbf{x})))}\Big) \nabla I^m(\mathbf{x} + \mathbf{u}(\mathbf{x})),\end{aligned} \tag{4}$$

where $\mathcal{X}_{\phi_i}(\mathbf{x}) = 1$ for all $\mathbf{x} \in \phi_i$ otherwise $\mathcal{X}_{\phi_i}(\mathbf{x}) = 0$. We require a priori distribution to be differentiable and strictly positive. Furthermore, coefficients β_i should be chosen to decrease the discrepancy in the magnitude of the computed force across the regions specially along the borders. We choose to solve the resultant nonlinear parabolic partial differential equation (PDE) using a semi-backward (implicit) Euler update formula as follows:

$$\mathbf{u}_d^{t+1} = (\mathbf{Id} + \alpha \delta t \mathbf{R})^{-1}(\mathbf{u}_d^t + \delta t \mathbf{f}_d(\mathbf{u}^t)), \tag{5}$$

where square matrix $\mathbf{R}$, with each dimension size equal to the total number of pixels in the image, represents the Laplacian operator Δ and $\mathbf{Id}$ is identity matrix

with the same size as **R**. Furthermore, d denotes the dimension (e.g., 1, 2, or 3), the superscript t is iteration step or time, and δt is the time marching coefficient [10]. The update equation 5 provides a greater stability and faster convergence compared to a simpler forward update formulation [7]. The downside is that it involves inverting a large matrix. To accelerate the inversion, we use a multi-grid method similar to the one outlined in [8].

4 Experimental Results

In this section, we explain the specific clinical application for the proposed method in details. Furthermore, we provide a detailed explanation regarding the input data and the desired output for this specific use case. Furthermore, we explain some implementation details and provide the validation results.

4.1 Adaptive Radiotherapy

During the planning phase in radiation oncology, tumors and the organs at risk are delineated and the dose distribution is planned using these contours. The patient undergoes a fractionated treatment process, where the incremental amount of dose is applied over several days. The original planned dose based on the planning CT is usually not valid throughout the whole treatment process, specifically for deformable targets such as prostate and organs at risk such as bladder and rectum with filling variations [6]. One solution is to acquire CT images prior to treatment in order to better capture the recent state of the anatomy. Two options are then available; first is to change the plan according to the recent state of anatomy, and second is keep the original plan but recompute the accumulated dose based on the current state of anatomy and to re-plan only if the accumulated dose deviates too much compared to the planned dose. Both of these options have pros and cons, however, both require an enabling technology that is to robustly find the corresponding structures (and voxels) between the planning and treatment scans. The main registration challenge here, as explained in the introduction section, is that *there are regions within these two images, where explicit correspondences cannot be established.*

4.2 Validation Results

We have acquired 20 pairs of CT data sets from 10 different patients undergone prostate radiotherapy. The first CT in each pair is the planning CT, where the contours of main three organs (prostate, bladder, and rectum) are available. The second CT in each pair is the treatment CT taken during one treatment session using a Siemens CTVision (Siemens Oncology Care Systems, Concord, CA, USA) within the treatment room. The three mentioned structures are then delineated in all the second CT data sets by expert dosimetrists. The regions of the images, which demonstrate significant intensity and shape variability without correspondence are considered to be the rectum, which is posterior to the

Table 1. Quantitative validation of the warped prostate, bladder, rectum contours as the results of the proposed method

Patient	Prostate			Bladder	Rectum
	ρ_d	ρ_{fd}	s_d (mm)	s_d (mm)	s_d (mm)
1	0.94	0.10	2.5	2.9	3.9
2	0.81	0.22	2.8	3.2	5.3
3	0.89	0.16	2.2	3.0	4.6
4	0.91	0.11	2.0	3.1	4.0
5	0.85	0.20	2.5	3.0	4.8
6	0.84	0.22	2.3	2.5	4.7
7	0.88	0.20	2.4	2.6	4.2
8	0.85	0.28	2.1	2.1	4.0
9	0.81	0.25	2.9	3.3	5.2
10	0.88	0.27	2.0	4.1	4.7
Average	0.87	0.20	2.4	3.0	4.1

Table 2. Quantitative validation of the warped prostate, bladder, rectum contours as the results of the conventional deformable registration algorithm with (a) SSD and (b) MI as (dis)similarity metrics

Patient	Prostate			Bladder	Rectum
	ρ_d	ρ_{fd}	s_d (mm)	s_d (mm)	s_d (mm)
1	0.66(a)/0.70(b)	0.28/0.15	3.9/4.0	5.1/5.2	6.4/5.9
2	0.71/0.54	0.38/0.34	4.0/4.0	4.9/5.2	5.9/6.1
3	0.62/0.75	0.40/0.21	4.0/3.7	4.9/5.1	7.5/8.2
4	0.73/0.69	0.39/0.46	4.1/4.1	4.9/4.8	6.7/7.1
5	0.66/0.52	0.41/0.23	3.8/4.0	5.3/5.0	7.2/5.9
6	0.69/0.53	0.38/0.37	3.7/4.0	5.0/5.0	7.0/6.9
7	0.59/0.61	0.43/0.34	4.3/4.2	5.1/4.7	7.6/6.9
8	0.68/0.62	0.19/0.22	4.5/4.1	5.2/4.8	8.0/7.8
9	0.68/0.59	0.31/0.29	4.9/3.8	4.7/5.2	6.8/6.1
10	0.67/0.71	0.20/0.31	4.3/4.2	4.6/4.9	7.1/6.8
Average	0.66/0.62	0.33/0.29	4.1/4.0	4.9/5.0	7.2/6.8

prostate and large intestine (bowel) that is superior to the bladder (see Figure 3). We used contours of the rectum in planning CT of first 10 pairs to estimate the intensity distribution of the rectum (shown in the Figure 4). We also used rough delineations of the large intestine region superior to the bladder of the first 10 pairs to estimate an intestine intensity distribution. We chose a rather large Parzen window of 35 on the Hounsfield scale to make sure the *pdf* is non-zero and smooth (differentiable) within the entire intensity range. We used the second 10 pairs of data as test data. We performed three sets of experiments. In the experiment set (a), we used the region based similarity metric, where the SSD is used for "regular" regions and the proposed similarity metric is used for the "special" regions delineated as the part of the planning process, in set (b), we used SSD as the intensity distance metric for the entire image, and finally

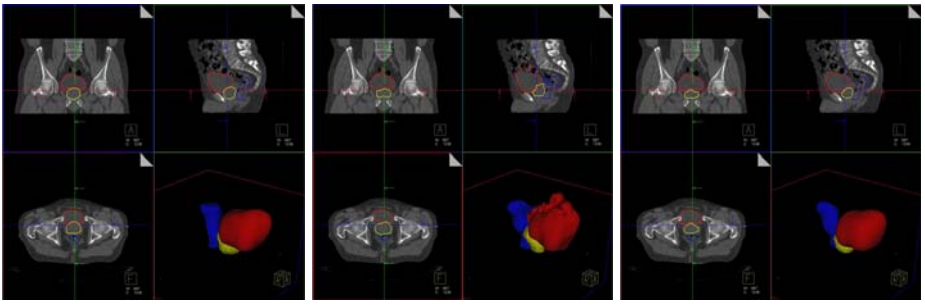


Fig. 3. The figures from left to right are, the treatment CT with ground truth contours of bladder, prostate and rectum, the treatment CT with warped version of the planning contours using a conventional method, and the treatment CT with warped version of the planning contours based on the proposed method

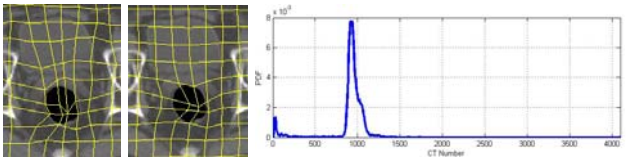


Fig. 4. Two figures on the left show the deformation grid overlaid on the planning data for *left* conventional method, and *right* proposed method. Right figure depicts the *pdf* of the rectum intensity (Hounsfield scale + 1024) based on 10 patient data.

in the set (c), we used MI. We always assume that planning CT is the fixed (reference) and the treatment CT is the moving (template) and we always first rigidly aligned the two images.

The images are all downsampled to $256 \times 256 \times \mathrm{N}$, where N is a number between 96 and 130. In order to cope with large displacements, equation 5 is solved in a scale-space setting. We solved the equations over three levels of resolutions starting with zero displacements as the initial value at the lowest resolution. For the three dimensional images of the mentioned size range, the overall time for registration process in average was less than 30 seconds. For these experiments, we empirically found appropriate values for α, β_0, and β_i once and kept them fixed for the entire test. Within the algorithm, the (dis)similarity metric force is always normalized by the initial (dis)similarity prior to the registration. This relieves us from adjusting the value of the α for various (dis)similarity metrics.

We used the inverse of the computed deformation field computed using the method described in [2] to generate the contours on the treatment data sets. For quantitative evaluation, we used several measures to compare the warped contours based on the registration process with the ground truth delineated by the expert clinician. The quantitative measures where; 1) ρ_d is the probability of detection, calculated as the fraction of the ground truth volume that overlaps with the estimated organ volume, 2)ρ_{fd} is the probability of false detection,

calculated as the fraction of the estimated organ that lies outside the ground truth organ, and 3) s_d is the surface distance, calculated as the average distance between the surfaces of the ground truth and estimated organs. We computed all measures for prostate but only s_d for bladder and rectum. In Tables 1 and 2, we show the measures computed for all three sets of experiments on the patient data. The proposed method always outperformed the conventional approaches, specially by higher margin for the cases where rectum shape changes were significant. Un-smooth deformation fields, which are typical results of the conventional method were problematic for the inverse deformation computation scheme. The examples of a specific patient data set is also brought in Figure 3.

5 Summary and Conclusion

In this paper, we presented a method for deformable registration of images, in a scenario, where there exist areas within the images without explicit correspondences. We take advantage of the assumption that although the pixel-to-pixel correspondence can-not be established, the intensity statistics within these areas remain consistent from image to image. Therefore, we proposed to use a softer region specific intensity matching or similarity constraint on these areas. This problem is clinically prevalent while processing various abdominal and pelvic scans of same patient at different time points. We outlined a detailed scheme for implementing the proposed method and focused on the problem of tumor and organ at risk variability in prostate radiotherapy and demonstrated the performance of the proposed solution.

References

1. Bajscy, R., Lieberson, R., Reivich, M.: A computerized system for the elastic matching of deformed radiographic images to idealized atlas images. J. Comput. Assis. Tomogr. 7, 618–625 (1983)
2. Chen, M., Lu, W., Chen, Q., Ruchala, K.J., Olivera, G.H.: A simple fixed-point approach to invert a deformation field. Med. Phys. 35(1), 81–88 (2008)
3. Christensen, G.E., Joshi, S.C., Miller, M.: volumetric transformation of brain anatomy. IEEE Trans. Medical Image 16, 864–877 (1997)
4. Ehrhardt, J., Schmidt-Richberg, A., Handels, H.: A Variational Approach for Combined Segmentation and Estimation of Respiratory Motion in Temporal Image Sequences. In: Proceedings of ICCV 2007, pp. 14–21 (2007)
5. Fischer, B., Modersitzki, J.: Fast diffusion registration. In: AMS Contemporary Math, Inverse Problems, Image Analysis, and Medical Imaging, vol. 313, pp. 117–129 (2002)
6. Foskey, M., Davis, B., Goyal, L., Chang, S., Chaney, E., Strehl, N., Tomei, S., Rosenman, J., Joshi, S.: Large deformation three-dimensional image registration in image-guided radiation therapy. Phys. Med. Biol. 50, 5869–5892 (2005)
7. Gonzales, C., et al.: Backward Euler Discretization of Fully Nonlinear Parabolic Problems. Mathematics of Computation 71(237), 125–145 (2001)

8. Henn, S., Witsch, K.: Iterative multigrid regularization techniques for image matching. SIAM J. Sci. Comput. 23(4), 1077–1093 (2001)
9. Maintz, J.B.A., Veirgerver, A.: A survey of medical image registration. Medical Image Analysis 2(1), 1–36 (1998)
10. Modersitzki, J.: Numerical Methods for Image Registration. Oxford University Press, New York (2004)
11. Rueckert, D., Frangi, A.F., Schnabel, J.A.: Automatic construction of 3D statistical deformation models using non-rigid registration. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 77–84. Springer, Heidelberg (2001)
12. Unal, G., Slabaugh, G.: Coupled PDEs for non-rigid registration and segmentation. In: Computer Vision and Pattern Recognition - CVPR 2005, vol. 1, pp. 168–175 (2005)

Intra-operative Multimodal Non-rigid Registration of the Liver for Navigated Tumor Ablation

Haytham Elhawary[1], Sota Oguro[1], Kemal Tuncali[2], Paul R. Morrison[2], Paul B. Shyn[2], Servet Tatli[2], Stuart G. Silverman[2], and Nobuhiko Hata[1]

[1] Surgical Planning Laboratory, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA, USA
{elhawary,sota,hata}@bwh.harvard.edu
[2] Abdominal Imaging and Intervention Group, Department of Radiology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA, USA
{ktuncali,pmorrison,pshyn,statli,sgsilverman}@partners.org

Abstract. CT guided tumor ablation of the liver often suffers from a lack of visualization of the target tumor and surrounding critical structures. This information is available on pre-operative contrast enhanced MR images and a non-rigid registration technique is desirable. However while registration methods have been successfully tested retrospectively on patient data, very few have been incorporated into clinical procedures. A non-rigid registration technique has been evaluated, optimized and validated to be able to perform registration of the liver between MR to CT images, and between intra-operative CT images. The method requires pre-processing and segmentation of the liver, and presents an accuracy of approximately 2mm. A clinical feasibility study has been conducted in 5 liver ablation cases. The method helps clinicians enhance interventional planning, confirm ablation probe location with respect to the tumor, and in the case of cryotherapy, evaluate tumor coverage by the ice ball.

1 Introduction

Liver cancer, namely hepatocellular carcinoma (HCC), is one of the most common cancers in the world, with an estimated 711,000 new cases reported worldwide in 2007 and 680,000 deaths [1]. Local tissue ablation has become a promising method of minimally invasive curative therapy that preserves uninvolved liver parenchyma with fewer systematic complications and side effects than major hepatic surgery [2]. Imaging technologies such as computed tomography (CT) can provide the guidance needed for the percutaneous placement of needle-like applicators for tumor ablation using chemical or thermal methods.

For procedure *planning*, an initial CT image of the liver is used to decide on an adequate entry point and probe tip location. The lack of soft tissue contrast of the CT image impedes a clear visualization of the tumor which is either barely distinguished or totally invisible. Most liver tumors are visible on pre-procedure contrast enhanced MRI and a visual correspondence of tumor location between the pre-procedure MRI and intra-operative CT images is performed by the physician using anatomical structures as landmarks. Once the position of the tumor is estimated and a plan devised, CT images are then taken to *guide* the probes into the planned location. Once the

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 837–844, 2009.

probes are positioned, CT images are taken to *monitor* the effects of the ablation, for example observing the extent of the ice ball in cryotherapy. The lack of proper visualization of the tumor can lead to sub-optimal ablation applicator placement and inadequate coverage of tumor margins. Clinically, this can increase complications and rates of recurrence [2].

Registration between the pre-procedure MR images and the acquired CT intra-operative images would allow mapping tumor location and size as well as any surrounding critical structures onto the CT images. Carrillo et al. [3] compared several semiautomatic voxel-based registration algorithms to perform rigid registration between pre-procedure MRI and interventional MRI of the liver. Niculescu et al. [4] registered pre- and post-treatment CT images by using a surface based non-rigid method with a finite element model to simulate volumetric deformation. Archip et al. [5] compared several non-rigid registration techniques to match pre-procedure contrast enhanced MR images of the liver to intra-operative CT, including a finite element based method they had previously developed for neurosurgery. However, while these registration methods have been successfully tested retrospectively on patient data, very few of these solutions have been incorporated into the clinical setting. In this paper we describe the optimization of a multimodal non-rigid registration technique which will be used to register pre-procedure contrast enhanced MR images to intra-operative CT images, and to register CT images obtained during different stages of the intervention. The technique is then incorporated in the clinical workflow of CT-guided liver tumor ablation, with a clinical feasibility study performed in 5 cases. This work will form part of a fully integrated liver ablation navigation suite we are currently developing.

2 Intra-operative Non-rigid Registration Method

Although rigid registration has been used before in CT guided liver ablation, the accuracy is not sufficient for guidance of therapy [6]. Several non-rigid registration techniques are reported in the literature and reviews of available methods have been extensively published [7]. Considering the trade off between computation requirements and accuracy [5] we decided to use a 3D volumetric non-rigid registration method which uses an affine transformation to model the global motion of the liver, and which describes local motion by a free-form deformation based on B-splines as described in [8] using the mutual information similarity metric. Three important factors must be considered for the adequate translation of a registration technique into the clinical setting: it must be *(i)* accurate for the clinical application at hand, *(ii)* fast enough to be performed within the time frame of a clinical intervention and *(iii)* robust so that it almost never fails. This section describes the optimization of the registration method and several pre-processing steps in order for it to fulfill these three conditions. The registration method along with all processing steps were made available in the open source software 3D Slicer [9].

Image Acquisition. The T1 weighted MR images were obtained using a 1.5T GE SIGNA scanner, 3D image acquisition (TR/TE=4.1/2ms, 5mm slice thickness, 2.5mm slice spacing, 512 x 512 matrix size, 400mm field of view) with an 8 channel torso

surface coil. The intra-operative CT images were obtained from a 40 channel Siemens Sensation Open, with a matrix size of 512 x 512.

Cropping, Segmentation and Slice Thickness. To eliminate irrelevant voxels in the registration process and to reduce the number of voxels in the computations, the image volumes can be cropped. The cropping could occur in plane or in the slice thickness direction. To improve registration accuracy the liver can be segmented from the images to produce masked data in order to include only the relevant pixels in the similarity measurement. All segmentation was performed by a qualified radiologist with 7 years experience. The CT images acquired during the interventional procedure could also be reconstructed at different slice thickness.

Registration Algorithm. Once all the pre-processing steps have been performed on the images, they are subject to the registration algorithm, which involves two steps. The first is to use an affine transformation to describe the global motion of the liver. The second step enforces local deformation by modeling the displacement field using a free-form deformation model based on B-spline interpolation, in order to maximize mutual information between the images. The efficiency of the registration algorithm depends on variables such as spatial samples, iterations and grid size, which will be modified to evaluate their effect on the registration.

2.1 Procedure Optimization

In order to study the influence of the various pre-processing steps and algorithmic variables on both the accuracy and the speed of registration, several experiments were performed to register 9 patient datasets. Table 1 shows the range of parameters that were considered. In total 9 different parameter values were evaluated over 9 patient datasets, giving a total of 81 registrations.

The first three parameter values consist of changing the slice thickness of the CT images to 3mm, 5mm and 7.5mm. Thinner slices provide increased detail although make the segmentation process longer. The next three set of parameter values considered were related with in-plane cropping. The in-plane field of view (FOV) was varied from being 100% of the acquired image, 66% of the image and then just a region surrounding the liver. The next parameter was manual segmentation of the liver, with the registration being performed with segmentation and then without to see its influence on accuracy. The following parameter was cropping in the slice thickness direction, with registration being done with over 75% of the liver volume included and then with between 50-75%. Finally the algorithm variables were modified in three categories, the first being high resolution (50,000 spatial samples for affine registration and 500,000 for B-spline), medium resolution (25,000 and 250,000) and low resolution (5,000 and 50,000). All the registration steps were performed on a PC with 4 Dual Core AMD Opteron 2.2GHz processors, 16GB of RAM with Fedore Core 10 Linux operating system.

For each set of parameters the time for the entire registration process including the pre-processing steps was measured. Following previous studies [5], quantifying registration accuracy was done by extracting the edges of the segmented livers from the registered datasets using a Canny Edge detector and the 95% Hausdorff distance (HD) calculated between the sets of points comprising the edges. Perfect alignment will produce a 95% Hausdorff distance of 0mm.

Table 1. Experiments performed to evaluate the effect of the different parameters on the registration process. Set 1 corresponds to a reference set of parameters, and each consecutive set involves modifying one parameter from the reference set.

Parameter Set	Slice Thickness (mm)	FOV Size	Segmentation	Liver Volume	Resolution
1 (Ref)	5	66%	Yes	>75%	High
2	3	66%	Yes	>75%	High
3	7.5	66%	Yes	>75%	High
4	5	100%	Yes	>75%	High
5	5	Liver	Yes	>75%	High
6	5	66%	No	>75%	High
7	5	66%	Yes	50-75%	High
8	5	66%	Yes	>75%	Med
9	5	66%	Yes	>75%	Low

2.2 Results

Fig. 1 shows the results of the registrations performed on the 9 patient datasets with the different parameters specified in Table 1. Regarding *accuracy* it can be seen that except for the case where the pre-processing has not included manual segmentation and masking of the data, average accuracy ranged from 1.9 - 3.4mm. This is a satisfactory result for liver tumor ablation, given that most treated tumors are large in size, ranging from 1-20cm. Without segmentation of the liver, registration showed large errors which average 10.6mm with a large standard deviation. Segmentation is therefore essential to the method. Regarding *time*, the entire registration process can take anything from 28-52 minutes when manual segmentation and masking is performed. Segmentation is by far the most time consuming step in the process taking between 10-15 minutes for each image. One way of reducing segmentation time is to increment slice thickness, as this means that there are fewer slices per volume. However at 7.5mm slice thickness the registration error starts to increase as does the standard deviation, and image resolution visibly decreases. Regarding *robustness*, the mutual information algorithm proves to be extraordinarily robust over the entire spectrum of parameter sets, except for the case with no segmentation. Small standard deviations mean that the accuracy of each individual registered dataset is close to the average, giving an indirect measure of robustness. For cropping both in-plane and in the slice direction the algorithm performed extremely well, hence incomplete image acquisitions of the liver should have little effect. After a thorough analysis of these results parameter set 1 was selected as the best protocol for use in the clinical procedure. All the pre-processing (cropping and segmentation) of the MR images can be done previous to the intervention, reducing the actual time needed during the clinical procedure to perform the registration.

With the parameters of the registration method defined, two further measures of accuracy were used to ensure validation before applying the protocol in the clinical procedure. The first was the Dice Similarity Coefficient (DSC) which gives a measure of the spatial overlap between the segmented liver from the registered MRI and

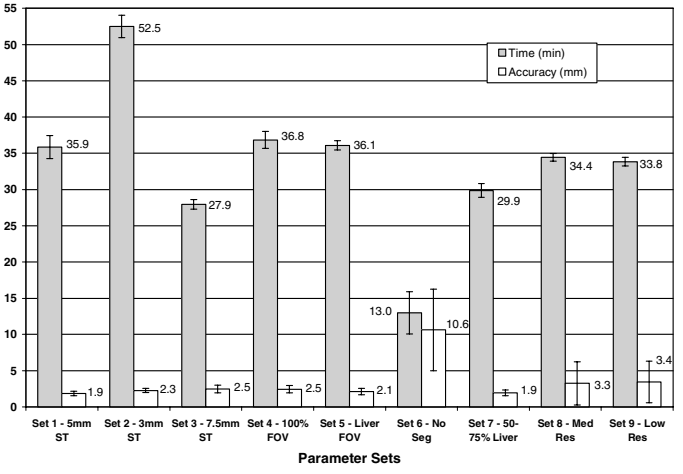


Fig. 1. Registration results for the experiments with different parameters. For each parameter set the average time and accuracy are presented, along with a standard deviation.

the CT image [10]. The DSC can range from [0,1], where 1 is perfect alignment. The second consists of asking an experienced radiologist to identify well defined corresponding features on the intra-operative CT image of the liver and on the registered MR image and to measure the 3D distance between them. Three anatomical points were selected in each dataset and the error for each of the three points was averaged. Table 2 shows the results obtained by performing both non-rigid and rigid registration on the 9 patient datasets with all three evaluation metrics. Using the non-rigid registration, an obtained improvement of four fold in the 95% HD, 13% in the DSC metric and 3.3 fold anatomical landmarks is presented.

3 Registration during CT Guided Tumor Ablation Procedure

The registration method was then integrated into the workflow of a CT guided liver tumor ablation procedure for a clinical feasibility study. It was applied in parallel during 5 cases of liver ablation, although no clinical decisions were made based on the information obtained from the registration, in accordance to approval by our Institutional Review Board. Table 3 shows the patients that underwent the intervention with non-rigid registration. Before the intervention, the pre-procedure MR images were cropped and manually segmented. Once the patient is located on the table of the CT scanner, usually in a different position than the supine in which the MRI is taken, a CT image is acquired to plan the intervention and sent to our workstation outside the scanner room for registration. The results of the registration are presented in the last two columns of Table 3, where both accuracy and time taken for the registration procedure during the intervention are displayed.

Table 2. Validation results for both rigid and non-rigid registration on 9 patient datasets using registration parameters set 1

Dataset	95% Hausdorff Distance (mm)		DSC		Anatomical Landmarks (mm)	
	Rigid	Non-rigid	Rigid	Non-rigid	Rigid	Non-rigid
1	9.00	1.41	0.86	0.97	8.15	1.21
2	4.58	1.41	0.83	0.96	11.30	3.19
3	15.94	2.24	0.79	0.96	23.01	3.77
4	8.60	2.02	0.85	0.96	10.51	6.11
5	6.48	2.23	0.84	0.96	8.81	5.05
6	3.74	1.73	0.92	0.96	4.04	2.73
7	6.16	1.97	0.87	0.97	5.33	2.24
8	9.00	2.01	0.84	0.97	15.67	2.57
9	8.12	1.96	0.90	0.98	12.55	3.28
Avg	**7.96**	**1.89**	**0.86**	**0.97**	**11.04**	**3.35**

Table 3. Patients receiving liver ablation procedure with registration, along with accuracy and total time of registration during procedure (not including the pre-processing of the MR images before the procedure)

Case	Sex	Age	Therapy	Pathology	Liver Segment	Accuracy (95% HD)	Time (min)
1	M	63	Cryotherapy	Metastasis	7	1.9	22
2	M	63	Cryotherapy	Metastasis	6/5	2.3	21
3	M	49	RF ablation	HCC	5	2	20
4	M	53	RF ablation	HCC	5	2.2	17
5	M	82	Cryotherapy	Metastasis	5	2.4	18

The registration takes approximately 17-22 minutes to complete, and during this time the surgeons plan the interventional procedure in their usual manner. When the registration is complete, the registered MR image is superimposed on the CT image and shown to the practitioners. The tumor is segmented from the MR image and overlaid on the CT image to delineate the exact target location (Fig. 2(a-c)).

Registration was also performed between the CT images obtained during the procedure. This process was done between the planning CT image taken at the beginning of the procedure, the CT image which displayed the ablation probes once positioned (before ablation started), and in the case of cryotherapy, the CT image showing the maximum sized ice ball. These images were registered to the original pre-procedural CT images using the registration method (segmentation was not required). Fig. 2(d)-(f) show the CT image after probe placement and the CT image after the maximum ice ball is performed, both overlaid with the segmented liver tumor taken from the MR image. Fig. 2(f) shows a 3D model of the liver, cryoablation probes, ice ball (all segmented from the CT image) and the tumor (segmented form the registered MR image). This type of model will allow a 3D evaluation of the tumor coverage and margins.

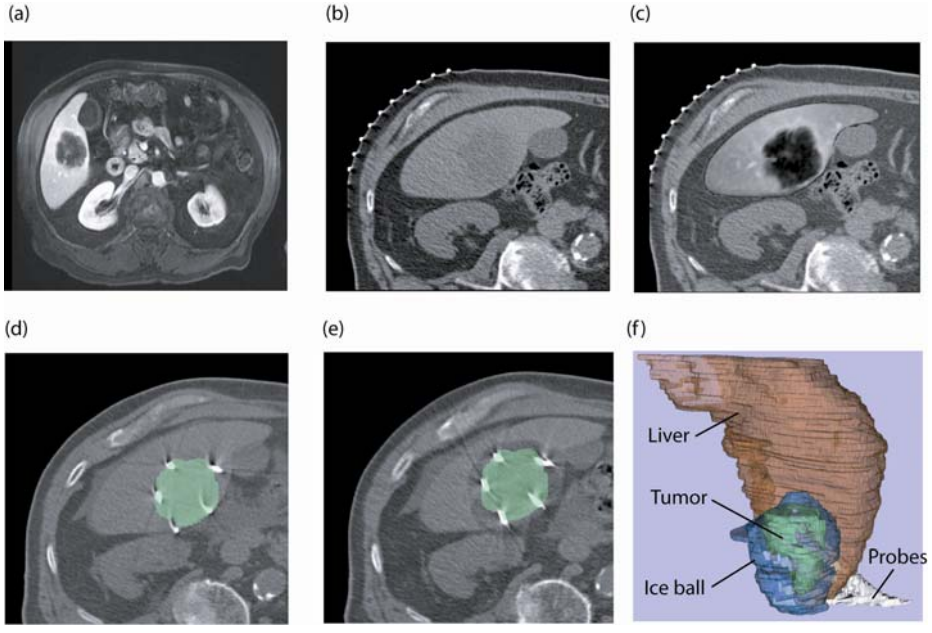


Fig. 2. (a) Pre-procedure contrast enhanced MR image, (b) CT image for interventional planning, (c) registered and fused image showing liver from MR image overlaid on CT image, (d) CT image showing ablation probes overlaid with segmented tumor from the registered MR image, (e) CT image showing iceball (hypointense area around lesion) overlaid with segmented lesion (green) from registered MR image, and (f) 3D representation of liver (red), tumor (green), iceball (blue) and cryoprobes to evaluate tumor coverage and margins

4 Discussion and Conclusion

The objective of this study was to provide an adequate registration algorithm for use during CT guided liver tumor ablation. This involves registering pre-procedure MRI images with intra-procedure CT images, and to register CT images taken at different phases of the procedure. Many of the registration methods described in literature are not applicable directly to this clinical problem because they require registration times that are unfeasibly long for their implementation during the clinical ablation procedure, and hence need optimization to find a compromise between accuracy and registration time. In this paper we have optimized and evaluated the accuracy, speed and robustness of a non-rigid voxel-based registration technique, which required preprocessing of the images including manual segmentation. Segmentation is the most time-consuming process of the whole registration method, and any gains would directly impact on the overall process and it's acceptance in the clinic. Future work is directed in evaluating semi-automatic segmentation algorithms, such as intensity based thresholding to produce an initial guess of the liver shape and an Estimation Maximization method [11]. We are also looking into computationally efficient implementations of Mutual Information [12]. The registration has been performed during 5 clinical cases, and with further shortening of registration times, will then be

applied in a prospective clinical study to evaluate the impact the registration has on clinical outcomes.

References

1. Garcia, M., Ward, E.M., Center, M.M., Hao, Y., Siegel, R.L., Thun, M.: Global Cancer Facts & Figures 2007. American Cancer Society, Atlanta (2007)
2. Tuncali, K., Morrison, P.R., Winalski, C.S., Carrino, J.A., Shankar, S., Ready, J.E., van Sonnenberg, E., Silverman, S.G.: MRI-guided percutaneous cryotherapy for soft-tissue and bone metastases: initial experience. AJR Am. J. Roentgenol. 189(1), 232–239 (2007)
3. Carrillo, A., Duerk, J.L., Lewin, J.S., Wilson, D.L.: Semiautomatic 3-D image registration as applied to interventional MRI liver cancer treatment. IEEE Transactions on Medical Imaging 19(3), 175–185 (2000)
4. Niculescu, G., Forand, D.J., Nosher, J.: Non-rigid Registration of the Liver in Consecutive CT Studies for Assessment of Tumor Response to Radiofrequency Ablation. In: Niculescu, G., Forand, D.J., Nosher, J. (eds.) 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007. EMBS (2007)
5. Archip, N., Tatli, S., Morrison, P., Jolesz, F., Warfield, S., Silverman, S.: Non-rigid Registration of Pre-procedural MR Images with Intra-procedural Unenhanced CT Images for Improved Targeting of Tumors During Liver Radiofrequency Ablations. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 969–977. Springer, Heidelberg (2007)
6. Wilson, D.L., Carrillo, A., Zheng, L., Genc, A., Duerk, J.L., Lewin, J.S.: Evaluation of 3D image registration as applied to MR-guided thermal treatment of liver cancer. Journal of Magnetic Resonance Imaging 8(1), 77–84 (1998)
7. Hill, D.L.G., Batchelor, P.G., Holden, M., Hawkes, D.J.: Medical image registration. Physics in Medicine and Biology 46(3), R1–R45 (2001)
8. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L.G., Leach, M.O., Hawkes, D.J.: Nonrigid registration using free-form deformations: application to breast MR images. IEEE Transactions on Medical Imaging 18(8), 712–721 (1999)
9. Gering, D.T., Nabavi, A., Kikinis, R., Hata, N., O'Donnell, J., Grimson, W.E.L., Jolesz, F.A., Black, P.M., W. III, W.M.: An Integrated Visualization System for Surgical Planning and Guidance Using Image Fusion and an Open MR. J. Magn. Reson. Imag. 13(6), 967–975 (2001)
10. Bharatha, A., Hirose, M., Hata, N., Warfield, S.K., Ferrant, M., Zou, K.H., Suarez-Santana, E., Ruiz-Alzola, J., D'Amico, A., Cormack, R.A., Kikinis, R., Jolesz, F.A., Tempany, C.M.: Evaluation of three-dimensional finite element-based deformable registration of pre- and intraoperative prostate imaging. Med. Phys. 28(12), 2551–2560 (2001)
11. Wells III, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. IEEE Transactions on Medical Imaging 15(4), 429–442 (1996)
12. Gholipour, A., Kehtarnavaz, N.D.: Computationally efficient mutual information estimation for non-rigid image registration. In: IEEE International Conference of Image Processing (2008)

A Novel Measure of Fractional Anisotropy Based on the Tensor Distribution Function

Liang Zhan[1,*], Alex D. Leow[1,2,*], Siwei Zhu[3], Marina Barysheva[1], Arthur W. Toga[1], Katie L. McMahon[4], Greig I. de Zubicaray[4], Margaret J. Wright[5], and Paul M. Thompson[1]

[1] Laboratory of Neuroimaging, Dept. of Neurology, University of California, Los Angeles, USA
[2] Department of Psychiatry, University of Illinois Medical Center at Chicago, USA
[3] Dept. of Mathematics, University of California, Los Angeles, USA
[4] Functional MRI Laboratory, Centre for Magnetic Resonance, University of Queensland, Brisbane, Australia
[5] Queensland Institute of Medical Research, Brisbane, Australia

Abstract. Fractional anisotropy (FA), a very widely used measure of fiber integrity based on diffusion tensor imaging (DTI), is a problematic concept as it is influenced by several quantities including the number of dominant fiber directions within each voxel, each fiber's anisotropy, and partial volume effects from neighboring gray matter. With High-angular resolution diffusion imaging (HARDI) and the tensor distribution function (TDF), one can reconstruct multiple underlying fibers per voxel and their individual anisotropy measures by representing the diffusion profile as a probabilistic mixture of tensors. We found that FA, when compared with TDF-derived anisotropy measures, correlates poorly with individual fiber anisotropy, and may sub-optimally detect disease processes that affect myelination. By contrast, mean diffusivity (MD) as defined in standard DTI appears to be more accurate. Overall, we argue that novel measures derived from the TDF approach may yield more sensitive and accurate information than DTI-derived measures.

1 Introduction

Diffusion-weighted MRI is a powerful tool to study water diffusion in tissue, providing vital information on white matter microstructure, such as fiber connectivity and integrity in the healthy and diseased brain. To date, most clinical studies still employ the *diffusion tensor imaging* (DTI) model [1,2], which describes the anisotropy of water diffusion in tissues by estimating, from a set of K diffusion-sensitized images, the 3x3 *diffusion tensor* (the covariance matrix of a 3-dimensional Gaussian distribution). Seven independent gradients are mathematically sufficient to determine the diffusion tensor, but MRI protocols with higher angular and radial resolutions, such as the high angular resolution diffusion imaging (HARDI) [3] or diffusion spectrum

* Zhan and Leow contributed equally (correspondences should be sent to Leow at feuillet@ucla.edu).

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 845–852, 2009.

imaging techniques [4], have been proposed to resolve more complex diffusion geometries, such as fiber crossings and intermixing of tracts. These geometries are incorrectly captured by a single-tensor model, as employed in standard DTI.

Among several recent advances in HARDI, the Q-ball imaging technique has been proposed to reconstruct fiber orientation density functions (ODFs) from the raw HARDI signal [3]. Deconvolution methods [5] can also yield mathematically rich models of fiber geometries using probabilistic mixtures of tensors [6], fields of von Mises-Fisher mixtures [7], or higher-order tensors (i.e., 3x3x…x3 tensors) [8,9]. Recent work on stochastic tractography [5, 10] also exploits the increased angular detail in HARDI. In most deconvolution-based methods, however, prior assumptions on fibers are usually imposed, e.g., all fiber tracts are forced to have the same anisotropy profile. Leow et al. recently proposed a more flexible approach, the Tensor Distribution Function (TDF) [11] to model fiber crossing in HARDI. Using the calculus of variations, the TDF approach can separate different dominant fiber directions in each voxel and compute their individual eigenvalues.

Much progress has been made in modeling more complex diffusion geometries that a single tensor fails to model, but most clinical studies still rely on simple DTI-derived scalar measures. Some of these, such as the trace of the covariance matrix or mean diffusivity (MD) can adequately describe isotropic water diffusion, but this only occurs in the cerebrospinal fluid spaces in the brain. In the white matter, myelinated fibers resist water diffusion orthogonal to the local dominant fiber orientation, and diffusion occurs preferentially along local fiber tracts. In clinical research, white matter fiber integrity is commonly assessed by determining how strongly diffusion is directionally constrained. One common scalar measure of directional diffusion, the fractional anisotropy (FA), is computed from the diffusion tensor's eigenvalues, and quantifies the magnitude of this directional preference. Clinical studies now routinely use FA as an index of white matter integrity, sensitive to white matter deterioration in aging and neurodegenerative diseases [12]. Even so, FA does not truly reflect the multidimensional complexity of the water diffusion profile. Regions with complex fiber-crossing tend to have lower FA values than predominantly unidirectional white matter structures (such as the midline corpus callosum; see Figure 2). However, it is unlikely that each of these crossing fibers in these regions has a true decrease in its integrity when compared to, say, corpus callosum fibers. In this paper, we argue that "white matter integrity", as measured by FA, is somewhat vague and imprecise, and may be greatly improved by using the full diffusion gradient information in HARDI. Factors that influence FA values may include the number of dominant fiber directions in each voxel, the eigenvalues of each of these fibers, partial volume effects from neighboring gray matter, and the non-Gaussianity of water diffusion. By using the TDF approach, which can separate crossing fibers, we examine where FA fails to reflect the underlying diffusion anisotropy.

2 Methods

Sixteen volunteers were scanned using a diffusion-sensitized MRI protocol on a Bruker Medspec 4 Tesla MRI scanner, with a transverse electromagnetic (TEM) headcoil. The timing and angular sampling of the diffusion sequence was optimized

for SNR [13]. The protocol used 94 diffusion-sensitized gradient directions, and 11 baseline scans with no diffusion sensitization (b-value: 1159 s/mm^2; TE/TR: 92.3/8250 ms; FOV=230x230; in-plane resolution: 1.8mmx1.8mm; 55 x 2mm contiguous slices; acquisition time: 14.5 minutes).

To process the data, a positive definite diffusion tensor was firstly estimated from the raw HARDI signal using the MedINRIA software (http://www-sop.inria.fr/asclepios/ software/MedINRIA), which projects the tensor manifold to its tangent plane at the origin to avoid negative or zero eigenvalues, which do not correspond to a physical diffusion process. Based on the diffusion tensor eigenvalues (λ_1, λ_2, and λ_3), the FA (defined in this paper as FA^{DTI} to avoid confusion) and MD (MD^{DTI}) may be calculated using Equation 1:

$$\begin{cases} FA^{DTI} = \sqrt{\frac{3}{2}\left(\frac{(\lambda_1-<\lambda>)^2+(\lambda_2-<\lambda>)^2+(\lambda_3-<\lambda>)^2}{\lambda_1^2+\lambda_2^2+\lambda_3^2}\right)} \\ MD^{DTI} =<\lambda>=\frac{\lambda_1+\lambda_2+\lambda_3}{3} \end{cases} \quad (1)$$

Values of FA^{DTI} range from 0 (no directional dependence of diffusion) to 1 (diffusion along a single direction). In addition to modeling the HARDI signal with a single tensor, we applied the framework in [11] to the HARDI signal, computing the voxelwise tensor distribution function P - a *pdf* defined on all physiologically feasible 3D Gaussian diffusion processes (in the tensor space **D)** at each voxel that best describes the observed signal. In our current implementation, we assume cylindrical fiber tracts and exclude planar-shaped tensors. Thus, $\lambda_1 \geq \lambda_2 = \lambda_3$ for each individual tensor in this tensor space. Given any TDF P, the number of dominant fibers is estimated by examining the local maxima of the tensor orientation distribution (TOD) $(TOD(\theta) = \int_{\lambda} P(D(\lambda,\theta))d\lambda)$. The eigenvalues of each fiber can be calculated by computing their expected values along the principal direction of this fiber. In this paper, we will investigate two sets of eigenvalues, i.e., those of the 1st dominant fiber $(\lambda_1^1, \lambda_2^1)$ and the voxelwise TDF-averaged eigenvalues $(\overline{\lambda_1}, \overline{\lambda_2})$ (Equation 2).

$$\begin{cases} \lambda_i^1 = \frac{\int P(D(\theta^*,\lambda))\lambda_i d\lambda}{\int P(D(\theta^*,\lambda))d\lambda} \quad (i = 1,2 \text{ and } \theta^* = argmax(TOD(\theta))) \\ \overline{\lambda_j} = \int_{D\in D} P(D)\lambda_j dD \quad (j = 1,2) \end{cases} \quad (2)$$

To obtain an overall assessment of the anisotropy in each voxel using the TDF framework, we propose the differential diffusivity (DD) $(DD = (\overline{\lambda_1} - \overline{\lambda_2})/(\overline{\lambda_1} + \overline{\lambda_2}))$. To compare TDF-derived measures to those from standard DTI, we compute the 1st dominant fiber's FA (FA^1) and MD (MD^1), as well as the TDF-averaged FA (FA^{TDF}) and MD (MD^{TDF}) by using the eigenvalues defined in Eq.(2) .

3 Results and Discussion

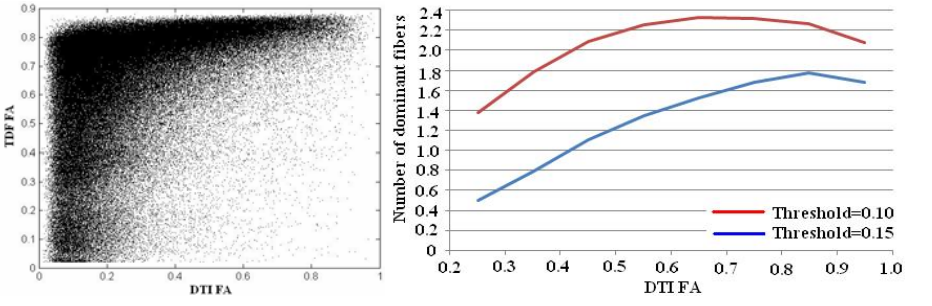
3.1 The Relation between DTI-FA and Actual Individual Fiber Anisotropy

We first assessed whether FA derived from standard DTI (FA^{DTI}) is an accurate measure of fiber anisotropy. Even though FA^{DTI} is usually intended to measure

white matter integrity, it in fact measures the compound effect of several factors, including the FA of individual fibers, the number of dominant fiber directions and partial volume effect from gray matter. Many of these will vary when the integrity of the fibers is not impaired, confounding its interpretation. Since any disease process that affects white matter myelination will most likely affect individual fiber anisotropy, we here restrict ourselves to investigating how well FA measures the anisotropy of individual fibers. As the TDF approach can separate multiple dominant fibers in one voxel and can determine their respective eigenvalues, we may answer the above question by investigating how well FA^{DTI} correlates with the FA

Table 1. Correlations between DTI-derived measures and TDF-derived measures for the whole brain and for voxels with FA^{DTI} >0.2 (thus mainly white matter voxels). FA^{DTI} correlates poorly with actual individual fiber anisotropy computed from the TDF. Compared to FA, MD^{DTI} correlates better with actual individual MD. All correlations reach statistical significance ($p<0.001$).

Correlations	FA^{DTI} *vs.* FA^{1}	FA^{DTI} *vs.* FA^{TDF}	FA^{TDF} *vs.* FA^{1}	MD^{DTI} vs. MD^{1}	MD^{DTI} vs. MD^{TDF}	MD^{TDF} vs. MD^{1}
Whole brain	0.309	0.433	0.884	0.546	0.620	0.849
Voxels with FA^{DTI} >0.2	0.155	0.285	0.827	0.440	0.509	0.821



(a) DTI FA *vs.* TDF FA (b) DTI FA *vs.* Number of dominant fibers

Fig. 1. Comparison between TDF and DTI. (a) Correlation between DTI FA and TDF FA. Here we plotted DTI-derived FA against TDF-averaged FA for the whole brain using one control subject as described in the methods section. FA derived from DTI is highly variable; if this measure does not accurately reflect the anisotropy of the component fibers, statistical power will be sacrificed when investigating group differences in white matter integrity using standard statistical testing (e.g., 2-sample *t* tests or multiple regressions). (b) DTI FA vs. Estimated number of dominant fibers. FA^{DTI} values are related to the number of dominant fiber directions, determined by thresholding the corresponding TOD at two different values (0.15 and 0.1) and counting the number of local maxima (e.g., at a threshold of 0.15, 61.2% of white matter voxels have one dominant fiber direction, 31.6% have two, and 7.2% have three or more). The mean number of dominant fiber directions for different ranges of FA^{DTI} is calculated by averaging the number of fiber directions within that range (e.g., when FA^{DTI} is within 0.6 and 0.7, the mean number of dominant fiber directions is 1.529). Both curves showed a non-linear relationship - with an initial increase for the number of fiber directions, followed by a decrease as FA^{DTI} increases.

values of the underlying dominant white matter fibers. As more than one fiber direction may be present, we examined how well FA^{DTI} correlates with either FA^{1} or the overall FA^{TDF}. Table 1 shows the results of these correlations. Only moderate correlations are found (0.471 for FA^{DTI} vs. FA^{TDF}, 0.352 for FA^{DTI} vs. FA^{1}) when both gray and white matter are included. Correlations are much weaker (0.303 and 0.164. respectively) when we only consider the white matter (operationally defined by a threshold of FA^{DTI} at 0.2), which is the tissue type FA was originally designed to investigate (FA^{TDF} correlates well with FA^{1}). By contrast, correlations are better between DTI- and TDF-derived MD measures (Table 1) than for FA (e.g., 0.631 for MD^{DTI} vs. MD^{TDF} in the whole brain). These results support our hypothesis that in the white matter, FA^{DTI} correlates very poorly with the actual individual fiber anisotropy, and thus (1) may be suboptimal for detecting subtle disease processes that affect myelination, and (2) may even be misleading, as low FA values may simply reflect the presence of multiple fibers.

To visualize these correlations, we plotted FA^{DTI} against FA^{TDF} in Figure 1(a). Visual inspection suggests an overall trend of positive correlation between these two measures, but a closer look suggests that FA^{DTI} is highly variable in the white matter. Indeed, for voxels with at least one dominant direction, the mean value of FA^{DTI} is 0.39 with a standard deviation (std) of 0.20, while those for FA^{TDF} are 0.74 and 0.16.

The observation that FA derived from DTI has a higher variability in the white matter is important for clinical applications where group comparisons are often conducted using a two-sample t test, or the general linear model. In these studies, the larger standard deviation of FA^{DTI} inevitably lowers the statistical power of a test with a fixed N (requiring more subjects to detect subtle changes). This further suggests that FA^{DTI} is a suboptimal measure of the underlying white matter myelination.

3.2 The Relation between FA^{DTI} and the Number of Dominant Fiber Directions

In the previous section, we showed that FA^{DTI} does not reflect the actual anisotropy of the underlying component fibers. In this section, we further explore the relation between FA^{DTI} and the number of dominant fibers. Here, we determine the number of dominant fibers by first thresholding the TOD at two different values (0.1 and 0.15) and counting the TOD local maxima exceeding these thresholds. As DTI is a single-tensor model, one would assume that the correlations with individual fiber measures would decrease (be less accurate) as the number of dominant fibers increases. This is indeed the case for MD (Table 2). However, the same analysis on FA (Table 2) did not reveal a similar trend. Instead, here the correlations first increase (with the number of dominant fibers) then decrease. To understand this discrepancy, we plotted the number of dominant fibers against FA^{DTI} in Figure 1(b), which reveals a similarly complex picture. Indeed, the number of dominant fibers first shows an increase when FA^{DTI} increases, followed by a decrease for voxels with the highest FA^{DTI} values. To understand this result, we hypothesize that the highest FA^{DTI} values usually correspond to predominantly uni-directional white matter structures (e.g., corpus callosum). These tend to be very heavily myelinated and have fewer dominant fiber directions compared to white matter voxels with fiber-crossing (thus lower FA^{DTI} values). However, at low FA^{DTI} values, we have to consider at least two opposite factors, both of which tend to cause a decrease in FA^{DTI} (but opposite trends for the number of

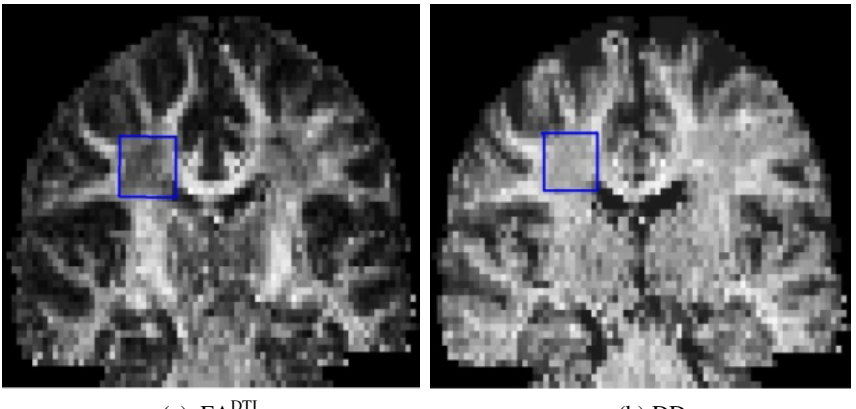
Table 2. Correlations between DTI-derived measures and TDF-derived measures for different numbers of dominant fibers

Correlations	MD^{DTI} *vs.* MD^{TDF}		FA^{DTI} vs. FA^{TDF}	
Dominant fibers	Whole brain	Voxels with $FA^{DTI}>0.2$	Whole brain	Voxels with $FA^{DTI}>0.2$
Number=0	0.532	0.523	0.269	0.131
Number=1	0.445	0.421	0.308	0.160
Number=2	0.377	0.361	0.314	0.213
Number≥3	0.317	0.307	0.298	0.211

fiber directions): the partial volume effect from gray matter (causing a *decrease* in the number of dominant fibers as the volume ratio of gray matter increases), and the fiber-crossing effect (causing an *increase* in number of dominant fibers as more fibers cross one another). The combination of these factors may explain the positive correlation of FA^{DTI} and the number of dominant fiber directions for low-to-medium FA^{DTI}. Our results are also consistent with those in [14], where FA was positively correlated with fiber density index (which measures how many fibers go through a given voxel) in patients with glioblastoma.

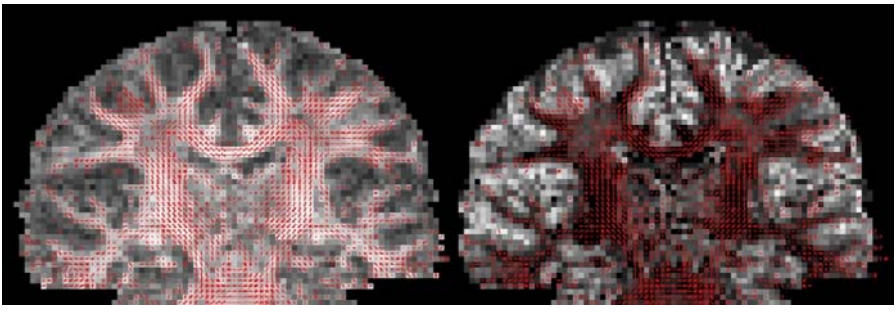
3.3 The Roles of Differential Diffusivity and Exponential Isotropy in Visualization

Here, we note that the TDF-derived differential diffusivity (DD), similar to FA, has higher intensities in the white matter than in the gray matter. However, unlike FA^{DTI},



(a) FA^{DTI} (b) DD

Fig. 2. FA^{DTI} is incorrectly depleted in regions with extensive fiber crossing (*blue box, inset*). By contrast, the differential diffusivity (DD), derived using the TDF approach, separates the dominant fiber directions and their corresponding anisotropy measures. In the highlighted region, fiber-crossing is present (fibers of the corpus callosum mix with the *corona radiata*, and superior longitudinal fasciculus). Here FA^{DTI} values are lowered relative to those of neighboring white matter, For DD, the signal is more consistent with that of the fibers entering the highlighted region.



(a)TOD overlaid with FA^{DTI} (b) TOD overlaid with EI

Fig. 3. Here, EI quantifies the overall isotropy of any given voxel, and highlights the gray matter instead of the white matter as in FA. EI, when overlaid on fiber directional information such as the TOD, may help visualizing HARDI data, offering information complementary to the FA map. Visually, the EI-TOD plot (right panel) provides a good representation of the spatial configurations of fiber tracts connecting with neighboring gray matter. Notice that in this figure, the TOD plot nicely resolves the fiber crossing in the Pons.

it does not suffer much from a drop in values in voxels with fiber-crossing (Figure 2). Figure 3 shows that, compared to FA, EI appears to help in visualizing fiber orientation mappings such as the TOD plot, as it complements the information in the TOD, and visually provides a good representation of the spatial configurations of fiber tracts connecting with the neighboring gray matter. Notice that in this figure, the TOD plot nicely resolves fiber-crossing in the Pons.

4 Conclusion

In this paper, we showed that white matter integrity, measured using the fractional anisotropy (FA) derived from standard DTI, is imprecise as it depends on several quantities including (among others) the number of dominant fiber directions, the anisotropy of each component fiber tract, and partial volume effects from neighboring gray matter. When compared with TDF-derived anisotropy measures, the FA obtained from standard DTI does not correlate well with the actual anisotropy of the individual component fibers, and may be sub-optimal in detecting subtle disease processes that affect white matter myelination. Future imaging studies of white matter integrity may benefit from assessing the number of dominant fiber directions in each voxel, and their corresponding eigenvalues and anisotropy. The TDF framework is ideal for achieving these goals. Lastly, we also demonstrated that the exponential isotropy (EI), differential diffusivity (DD) and the tensor orientation distribution (TOD) (all novel concepts derived from TDF) may help in visualizing HARDI data, as they provide additional information complementary to measures obtained from standard DTI.

References

1. Basser, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by quantitative diffusion tensor MRI. J. Magn. Reson. B 111(3), 209–219 (1996)
2. Le Bihan, D.: IVIM method measures diffusion and perfusion. Diagn Imaging (San Franc) 12(6), 133–136 (1990)
3. Tuch, D.S.: Q-Ball Imaging. Magnetic Resonance in Medicine 52, 1358–1372 (2004)
4. Tuch, D., Diffusion, M.R.I.: of complex tissue structure. PhD thesis, Harvard University-Massachusetts Institute of Technology, Cambridge, Massachusetts (2002)
5. Alexander, D.: Maximum entropy spherical deconvolution for diffusion MRI. In: Proceedings of the 19th International Conference on Information Processing in Medical Imaging (IPMI), Glenwood Springs, CO, USA (2005)
6. Hess, C.P., Mukherjee, P., Han, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. Magn. Reson. Med. 56, 104–117 (2006)
7. Jansons, K.M., Alexander, D.: Persistent angular structure: new insights from diffusion magnetic resonance imaging data. Inverse Probl. 19, 1031–1046 (2003)
8. Anderson, A.: Measurement of fiber orientation distributions using high angular resolution diffusion imaging. Magn. Reson. Med. 54, 1194–1206 (2005)
9. Özarslan, E., Shepherd, T., Vemuri, B.C., Blackband, S., Mareci, T.: Resolution of complex tissue microarchitecture using the diffusion orientation transform (DOT). NeuroImage 31(3), 1083–1106 (2006)
10. Tournier, J.D., Calamante, F., Gadian, D., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage 23, 1176–1185 (2004)
11. Leow, A.D., Zhu, S., Zhan, L., de Zubicaray, G.I., Meredith, M., Wright, M., Toga, A.W., Thompson, P.M.: The Tensor Distribution Function. Magn Reson Med. 18; 61(1), 205–214 (2009)
12. Zhang, Y., Schuff, N., Jahng, G.H., Bayne, W., Mori, S., Schad, L., Mueller, S., Du, A.T., Kramer, J.H., Yaffe, K., Chui, H., Jagust, W.J., Miller, B.L., Weiner, M.: Diffusion tensor imaging of cingulum fibers in mild cognitive impairment and Alzheimer disease. Neurology 68(1), 13–19 (2007)
13. Jones, D.K., Horsfield, M.A., Simmons, A.: Optimal strategies for measuring diffusion in anisotropic systems by magnetic resonance Imaging. Magn. Res. Med. 42(3), 515–525 (1999)
14. Roberts, T.P.L., Liu, F., Kassner, A., Mori, S., Guha, A.: Fiber Density Index Correlates with Reduced Fractional Anisotropy in White Matter of Patients with Glioblastoma. AJNR Am J. Neuroradiol. 26, 2183–2186 (2005)

Iterative Co-linearity Filtering and Parameterization of Fiber Tracts in the Entire Cingulum

Marius de Groot[1], Meike W. Vernooij[2,3], Stefan Klein[1], Alexander Leemans[4,5], Renske de Boer[1,2], Aad van der Lugt[3], Monique M.B. Breteler[2], and Wiro J. Niessen[1,6]

[1] Biomedical Imaging Group Rotterdam, Departments of Radiology and Medical Informatics, Erasmus MC, Rotterdam, the Netherlands
marius.degroot@erasmusmc.nl
[2] Department of Epidemiology, Erasmus MC, Rotterdam, the Netherlands
[3] Department of Radiology, Erasmus MC, Rotterdam, the Netherlands
[4] Image Sciences Institute, University Medical Center Utrecht, the Netherlands
[5] CUBRIC, School of Psychology, Cardiff University, United Kingdom
[6] Imaging Science and Technology, Faculty of Applied Sciences, Delft University of Technology, the Netherlands

Abstract. We present a method for the fully automated extraction of the cingulum using diffusion tensor imaging (DTI) data. We perform whole-brain tractography and initialize tract selection in the cingulum with a registered DTI atlas. Tracts are parameterized from which tract co-linearity is derived. The tract set, filtered on the basis of co-linearity with the cingulum shape, yields an improved segmentation of the cingulum and is subsequently optimized in an iterative fashion to further improve the tract selection. We evaluate the method using a large DTI database of 500 subjects from the general population and show robust extraction of tracts in the entire cingulate bundle in both hemispheres. We demonstrate the use of the extracted fiber-tracts to compare left and right cingulate bundles. Our asymmetry analysis shows a higher fractional anisotropy in the left anterior part of the cingulum compared to the right side, and the opposite effect in the posterior part.

1 Introduction

The use of diffusion tensor magnetic resonance imaging (DTI) for analyzing brain microstructure is well established [1]. The diffusion tensor, estimated from multiple diffusion weighted image volumes, captures the local organization of the brain. The fractional anisotropy (FA), derived from this tensor, describes the microstructural tissue organization in terms of the degree of diffusion directionality [2]. Comparing differences between neuroanatomical structures across subjects requires one to establish correspondence between these subjects. This may involve a common white matter skeleton [3], or the identification of separating manifolds between white matter structures [4] for the projection of individual measures onto a reference frame. Compared to FA extraction, streamline tractography in the tensor field provides a richer description of the local white matter structure. Tractography based comparison between subjects will therefore be more sensitive to subtle differences, provided, of course, that the abundance of data in the tracts can be interpreted.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 853–860, 2009.

In this paper, we develop a method to automatically analyze the cingulum in a large cohort using streamline tractography.

The Cingulum. The cingulum is a tube-like structure that courses from the hippocampal formation and bends around the ventricles and the corpus callosum (cc) up to the genu of the cc. The anterior cingulate is thought to be primarily involved in executive functioning, whereas the posterior cingulate is assumed to play a role in memory processes. Its role in cognitive functioning has generated a large interest in the analysis of various aspects of the cingulum. Due to high curvature in its local anatomy, the cingulum is difficult to reconstruct with streamline tractography past the bending point around the splenium of the cc [5]. This may explain why most previous research focused on tracking only the anterior part of the cingulum. Research aiming to study the cingulum with the use of tractography in multiple subjects, has often relied on manual ROI definition to allow tract selection (e.g. [6-8]). O'Donnell et al. use a spectral embedding function to identify tracts similar to a template tract [9]. Their 'tract based morphometry' can robustly capture tracts in the anterior cingulate bundle, but has not been able to identify tracts when every tract only overlaps partly with the anatomical structure of interest. Depending on DTI data quality, the high curvature in the cingulum may pose a particular challenge for tractography. We propose an automated method that does not rely on individual tracts coinciding with the entire cingulum, but rather on extracting the parts of all tracts that describe at least some part of the cingulum. The complete filtered set of all (partial) tracts then has more power to capture the actual cingulum anatomy, in particular the posterior part.

2 Methods

In short, after performing whole-brain tractography, we propose to identify partial tracts that make up the cingulate bundle with the use of a registered anatomical atlas. We derive a parameterization, which allows for a more accurate selection of tracts that belong to the cingulum and facilitates inter-subject comparison. The iterative approach, outlined in Fig. 1, consists of the following steps: tract selection, derivation of the parameterization, projection of tracts and co-linearity filtering. In the methods description, all parameters that need to be set are printed in italic.

Tractography and Atlas Registration. Deterministic streamline tractography is performed using the 'ExploreDTI' package (www.exploredti.com) [10]. Tracts are described by cubic B-splines, and stored as equally spaced points. We adopted the neuroanatomical ICBM-DTI-81 atlas of Mori et al. [11] for the initial tract selection. This atlas is created by manual segmentation of 50 neuroanatomical structures. We nonrigidly registered the atlas' FA image towards all subjects' FA image using 'elastix', a publicly available software package that uses an efficient stochastic gradient descent optimization method [12]. The registration is used to transform the atlas' labels to each subject's native space, resulting in an initial three-dimensional segmentation of the cingulate bundle. In following iterations, an improved segmentation is used, which is derived after filtering for co-linearity.

Tract Selection. The cingulate template is used to identify tracts running inside the cingulum. Owing to the absence of nearby co-linear structures, we assume that fiber tracts which are partially within the cingulum before they divert due to, e.g., partial voluming effects, can be included in the analysis [9]. Tracts are cut off when leaving the segmented region unless they reenter the segmentation within a maximum distance (*maxDist*). In this case they are considered part of the cingulum and included in the filtered tract set. A tract-density map is generated from the identified tracts at a 1mm cubic grid, indicating the number of cingulum tracts visiting each voxel.

Derivation of the Centerline and Projection of Tracts. By extracting the centerline of the cingulum, we establish a parameterization axis for the tract set. We achieve this by treating the tract-density map as a reciprocal cost image, and find the minimum cost centerline from one end of the cingulum to the other. To bridge potential gaps in the tract sets, we allow background voxels to be part of the centerline, but only at high costs (*backgroundWeight*). The parameterization should permit direct inter-subject comparison and should therefore not depend on individual performance of the fiber tracking. Especially at the far ends of the cingulum, tracts are often hard to reconstruct, leading to decreased tract density, which in turn could lead to problems in establishing correspondence across subjects. We therefore add the registered atlas mask to the tract density map with a fixed weight (*atlasWeight*), prior to converting it to the cost image. In dense regions, this has little effect, but in sparse regions, this allows to find the centerline aided by the registered atlas. The centerline is smoothed by fitting a B-spline with a fixed number of control points (*controlPoints*). The parameterization is then super sampled using *parSegments* equally spaced segments.

We then use the planes normal to the parameterized centerline to associate every tract segment to one of the parameterization segments.

Co-linearity Filtering. Because of the linear structure of the cingulum, it is safe to assume tracts running inside the cingulum are to a certain extent co-linear. Therefore, for every tract segment, we use the local direction to calculate co-linearity defined as the absolute dot product of the tract and parameterization segment direction vectors. A filtering step is then applied to exclude tract segments that are not co-linear (less than *minColin*) with the parameterization for a prolonged segment length (*maxDist*). After filtering, we construct a binary segmentation of all voxels containing tract(s).

Iterative Approach. The binary tract mask can be regarded as an improved segmentation of the cingulate bundle. It has been allowed to expand, by tracts locally exiting and re-entering the previous segmentation. And it has been allowed to shrink, by the local absence of co-linear tract segments. The mask is used as an updated segmentation for the entire procedure. The whole procedure is performed iteratively, until either follow-up iteration masks disagree on less than a fixed number of voxels (*maxVoxels*), or when a maximum number of iterations (*maxIt*) has been performed.

Statistical Analysis. Measures describing tracts and tract locations, such as FA, tract curvature etc, are at this point associated with a single linear parameterization and can thus be compared across subjects. In our application study, we chose to study hemispheric asymmetry in FA of the cingulum. The FA is a well-established measure and allows comparison of our results with previously published findings.

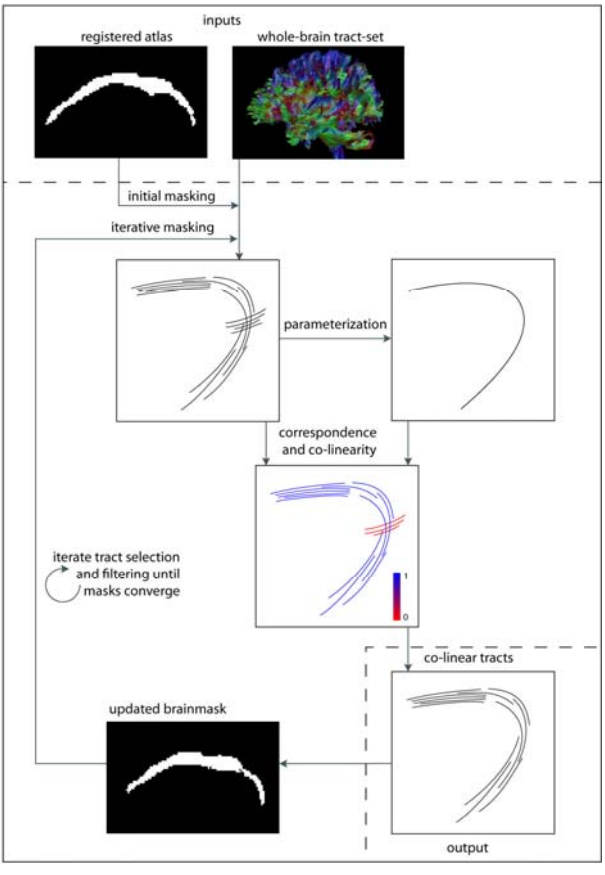


Fig. 1. Schematic overview of the tract selection procedure

3 Experiments and Results

DTI data from the population based Rotterdam Scan Study is used [13]. From this prospective cohort study, a random subset of 500 datasets was selected. Mean age of participants was 54.9 years, with an SD of 5.52; 52 percent of participants was female. All subjects were scanned on a 1.5 Tesla GE MRI system, with the following DTI acquisition parameters: 25 diffusion gradient directions, FOV= 210x210 mm^2, scan matrix= 96x64 (zero padded in k-space to 256x256), slice thickness 3.5 mm, 35 contiguous slices, TR= 8000 ms, TE= 68.7 ms, b-value= 1000 s/mm^2 and 3 volumes acquired without diffusion weighting. Datasets were preprocessed using the FDT toolbox in FSL [14] to correct for head motion and Eddy currents and to calculate FA images for the atlas registration. Atlas registration was performed using a 10 mm B-spline control point spacing with mutual information as similarity measure and took 8 minutes per subject on a single standard CPU. Streamline tractography was performed seeding tracts in a 2 mm cubic grid; and ending tracts when FA dropped below 0.2, taking 2 minutes per subject to compute.

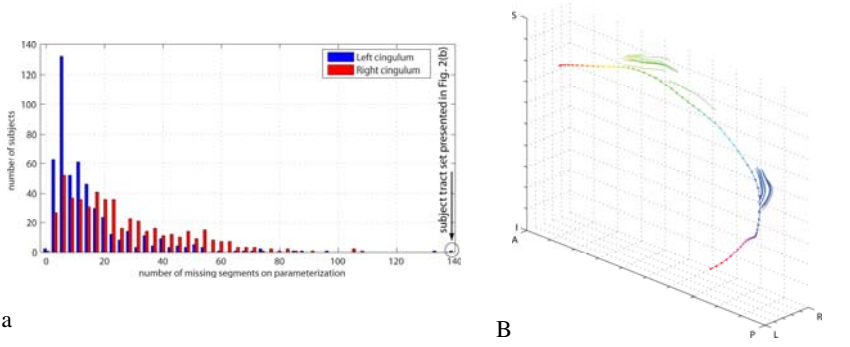


Fig. 2. (a) Histogram showing the number of missing segments per subject. (b) Most problematic tract set identified in histogram. The dotted line indicates the centerline; parameterization position is indicated by color, parameterization legend in Fig. 3(b).

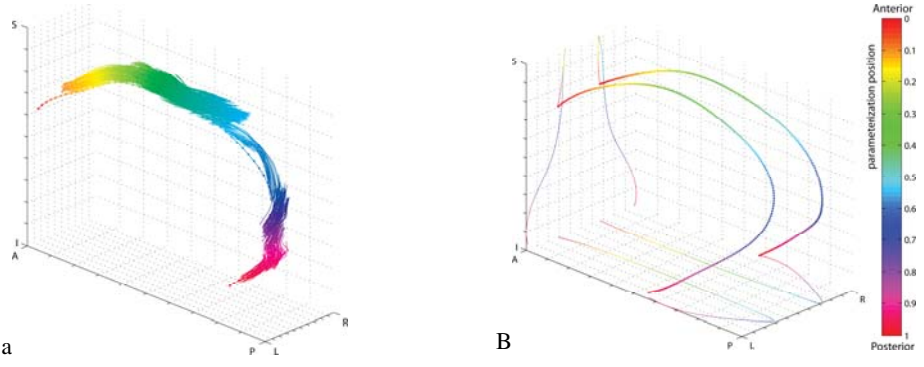


Fig. 3. (a)Typical tract set. The dotted line indicates the centerline; parameterization position is indicated by color. (b) Mean parameterization of the left and right cingulate bundles, serving as parameterization legend for Fig. 2(b), 3(a), 4 and 5. Thin lines indicate projections.

The applied setpoints have been determined empirically by visual inspection of resulting tract sets. The *maxDist* length of a tract part that is allowed outside the segmentation, as well as the maximum non-collinear length has been set to 3mm. The *backgroundWeight*, allowing the parameterization to cross over gaps, is set to 5 times the maximum costs encountered in the individual cost image. The *atlasWeight* is set to 1. The number of *controlPoints* of the parameterization is 15, later super sampled to 200 *parSegments*. Tract segments need to have a *minColin* of 0.8 to be considered co-linear with the parameterization. Iteration stops once the tract mask changes at most *maxVoxels*, 10, in a single iteration, or after reaching *maxIt*, 5, iterations. Iterative filtering took 9 minutes to compute; per subject per hemisphere.

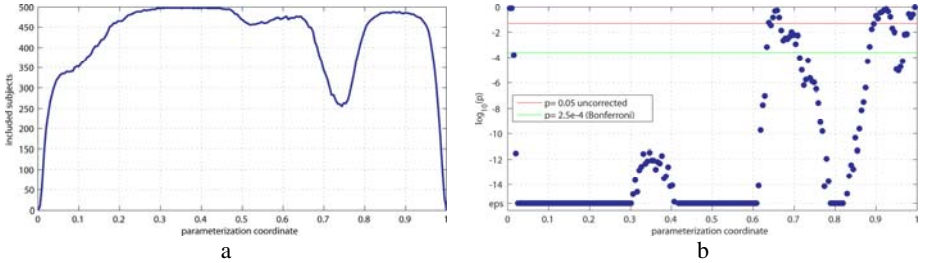
The tract selection method does not require a tract to be mapped at every parameterization segment. Fig. 2(a) shows a histogram of the number of associated parameterization segments per subject. Ideally, all subjects should have no missing segments, this is however not the case. For the worst performing subject, the rightmost element in Fig. 2(a), the tract set is shown in Fig. 2(b). The procedure identified a very limited

number of tract segments; yet it did not make any apparent mistakes in corresponding the tract set. A typical tract set is shown in Fig. 3(a).

The FA is compared between the left and right cingulum for every segment of the parameterization, using a two-tailed paired t-test. Subjects therefore need to have a mean FA value associated with a particular segment on both their left and right cingulum. Failure in either or both hemispheres leads to exclusion of that subject from the analysis at that particular parameterization segment. Fig. 4(a) shows the number of subjects included for every tract segment. Resulting log(p)-values, shown in Fig. 4(b), show highly significant differences in both the anterior and posterior part of the cingulum. Mean parameterization coordinates are presented in Fig. 3(b) for reference, the mean FA along the left and right cingulate bundles is presented in Fig. 5(a).

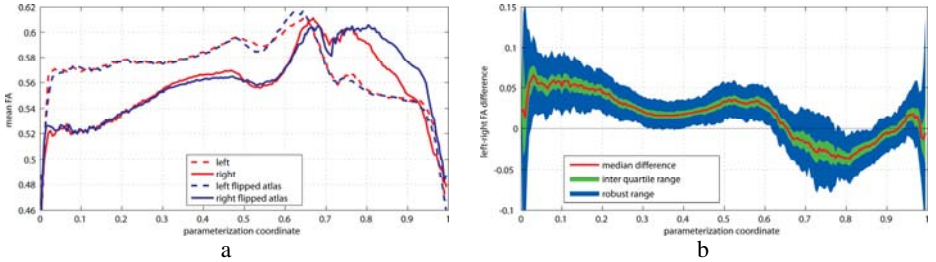
As we are studying hemispheric asymmetry, it is important to rule out asymmetric influence of the atlas. To verify our findings, we flipped the atlas over the sagittal plane and repeated the registrations to the subjects' FA images. These new registrations were used to again measure FA profiles in cingulate bundles for all subjects. The blue lines in Fig. 5(a) present the mean FA profiles for left and right cingulate bundles, extracted with the flipped atlas, where left means subject-left. The profiles appear nearly identical to the original results, ruling out a bias by the atlas.

To investigate robustness of our findings, we evaluated the mean left – right FA in a subset of 50 subjects, repeatedly drawn random from the full set; shown in Fig. 5(b).



a b

Fig. 4. (a) Number of complete cases, i.e. subjects with FA measurements in both left and right cingulate bundle, tested per parameterization segment. (b) Significance per tract segment, reported in log(p)-values.



a b

Fig. 5. (a) Mean FA in left and right cingulate bundles, compared to mean FA obtained using a flipped atlas. (b) Robustness of left-right FA difference, measured by 200 random draws of 50 subjects from the total set of 500 subjects.

4 Discussion

We have developed a novel method for selecting tracts of interest from a whole brain tract set. The approach is especially useful in large datasets and is aimed at investigating tracts that are difficult to follow with more conventional region of interest tract selection approaches. The present work allows tracking of tube-like structures such as the cingulate bundle or the fornix. We acknowledge that the method contains a number of thresholds, albeit most of them having a physical, intuitive meaning, which have to be predefined. These thresholds were determined by qualitative evaluation of the resulting tract selections. To validate our method we compare our analysis results with published findings. To this extent we have studied inter-hemispheric differences in cingulate FA using a large sample of 500 subjects from the general population. We have shown that our findings are not caused by a bias in the initial atlas segmentation. In the anterior part of the cingulum, we observe a very significant difference with the FA being higher left than right. This is in agreement with results from other analyses that also used some form of tract selection [6, 7, 9]. Less is known, however, for the posterior cingulate. Malykhin et al. [7] selected tracts in this region in 24 healthy subjects. They found no significant difference between left and right tracts, possibly lacking statistical power. As the anterior and posterior cingulate are involved in different processes, the left/right difference in FA between the anterior and posterior cingulum might be hypothesized to reflect a difference in function between left and right hemispheric processes.

We tested for differences using a paired t-test on the parameterization-segment level, and reported p-values for every segment. To correct p-values for the number of tests performed (we tested on 200 segments), a permutation based correction method would have been appropriate [15]. Applying a Bonferroni correction results in an overly conservative upper bound to the multiple comparison correction, but still leaves our findings highly significant, as indicated by the green threshold in Fig. 4(b).

The method allows capturing the difficult connection between the posterior and anterior cingulate, but it does so with varying success along the pathway. For the most problematic segment, about half of the subjects give problems in either or both of the cingulate bundles, as shown in Fig. 4(a).

5 Conclusion

We have developed a fully automated method for the robust extraction and parameterization of the cingulum including the posterior cingulate. As an example application, we studied asymmetry of FA in the cingulum. Our findings in the anterior part, where FA was found to be higher left than right, are in agreement with previous research. Our analysis also found right higher than left FA in the posterior part.

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophysical journal 66, 259–267 (1994)
2. Basser, P.J., Jones, D.K.: Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. NMR in Biomedicine 15, 456–467 (2002)

3. Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. NeuroImage 31, 1487–1505 (2006)
4. Kindlmann, G., Tricoche, X., Westin, C.F.: Delineating white matter structure in diffusion tensor MRI with anisotropy creases. Medical image analysis 11, 492–502 (2007)
5. Melonakos, J., Mohan, V., Niethammer, M., Smith, K., Kubicki, M., Tannenbaum, A.: Finsler Tractography for White Matter Connectivity Analysis of the Cingulum Bundle. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 36–43. Springer, Heidelberg (2007)
6. Gong, G., Jiang, T., Zhu, C., Zang, Y., Wang, F., Xie, S., Xiao, J., Guo, X.: Asymmetry analysis of cingulum based on scale-invariant parameterization by diffusion tensor imaging. Human brain mapping 24, 92–98 (2005)
7. Malykhin, N., Concha, L., Seres, P., Beaulieu, C., Coupland, N.J.: Diffusion tensor imaging tractography and reliability analysis for limbic and paralimbic white matter tracts. Psychiatry research 164, 132–142 (2008)
8. Nakata, Y., Sato, N., Nemoto, K., Abe, O., Shikakura, S., Arima, K., Furuta, N., Uno, M., Hirai, S., Masutani, Y., Ohtomo, K., Barkovich, A.J., Aoki, S.: Diffusion abnormality in the posterior cingulum and hippocampal volume: correlation with disease progression in Alzheimer's disease. Magnetic resonance imaging 27, 347–354 (2008)
9. O'Donnell, L.J., Westin, C.F., Golby, A.J.: Tract-based morphometry for white matter group analysis. NeuroImage 45, 832–844 (2009)
10. Leemans, A., Jeurissen, B., Sijbers, J., Jones, D.K.: ExploreDTI: a graphical toolbox for processing analyzing, and visualizing diffusion MR data. In: 17th Annual Meeting of Intl. Soc. Mag. Reson. Med., Hawaii, USA, p. 3537 (2009)
11. Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A.V., Mahmood, A., Woods, R., Toga, A.W., Pike, G.B., Neto, P.R., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P., Mazziotta, J.: Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. NeuroImage 40, 570–582 (2008)
12. Klein, S., Staring, M., Pluim, J.P.W.: Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines. IEEE Trans. Image. Process 16, 2879–2890 (2007)
13. Hofman, A., Breteler, M.M., van Duijn, C.M., Krestin, G.P., Pols, H.A., Stricker, B.H., Tiemeier, H., Uitterlinden, A.G., Vingerling, J.R., Witteman, J.C.: The Rotterdam Study: objectives and design update. European journal of epidemiology 22, 819–829 (2007)
14. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M.: Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23 (suppl. 1), S208–S219 (2004)
15. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: a primer with examples. Human brain mapping 15, 1–25 (2002)

Think Global, Act Local; Projectome Estimation with BlueMatter*

Anthony J. Sherbondy[1], Robert F. Dougherty[2], Rajagopal Ananthanarayanan[1], Dharmendra S. Modha[1], and Brian A. Wandell[2]

[1] IBM Almaden Reserach Center, Almaden, USA
anthony.sherbondy@stanford.edu
[2] Psychology Department, Stanford University, USA

Abstract. Estimating the complete set of white matter fascicles (the projectome) from diffusion data requires evaluating an enormous number of potential pathways; consequently, most algorithms use computationally efficient greedy methods to search for pathways. The limitation of this approach is that critical global parameters - such as data prediction error and white matter volume conservation - are not taken into account. We describe BlueMatter, a parallel algorithm for global projectome evaluation, which uniquely accounts for global prediction error and volume conservation. Leveraging the BlueGene/L supercomputing architecture, BlueMatter explores a massive database of 180 billion candidate fascicles. The candidates are derived from several sources, including atlases and mutliple tractography algorithms. Using BlueMatter we created the highest resolution, volume-conserved projectome of the human brain.

1 Introduction

The white matter of the human brain comprises more than $150km$ of long-range myelinated connections [1]. Understanding the architecture of these long-range projections (the projectome) is important for understanding brain function [2]. Diffusion weighted imaging fiber tractography (DWI-FT) is the only non-invasive method for studying the human brain projectome.

While there has been great progress in developing fiber tracking techniques [3,4,5], there is wide agreement that current methods fail in many specific cases [6,7,4]. A limitation is that current algorithms find pathways using greedy techniques; that is, the algorithms make decisions based on individual tracts without considering the entire projectome. Further, current algorithms do not optimize

* Sponsored by Defense Advanced Research Projects Agency, Defense Sciences Office (DSO), Program: Systems of Neuromorphic Adaptive Plastic Scalable Electronics (SyNAPSE), Issued by DARPA/CMO under Contract No. HR0011-09-C-0002. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressly or implied, of the Defense Advanced Research Projects Agency or the U.S. Government. Also supported by NEI EY01500.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 861–868, 2009.

the projectome to fit the original diffusion data. Finally, they ignore important physical constraints, such as the volume occupied by the estimated fascicles.

We introduce the BlueMatter algorithm to address these limitations. BlueMatter takes as input fascicles derived from multiple tractography algorithms. It searches for an optimal combination of these fascicles (the projectome) subject to two error terms. The first term compares the predicted and measured diffusion-weighted images. We refer to this constraint as diffusion-fitting (see also [8,9]). The second term accounts for an upper limit on the fascicle volume. We refer to this constraint as volume regularization; this term is an important physical constraint that helps resolve the ill-posed inverse problem of finding a projectome that minimizes diffusion-fitting error.

Accounting for diffusion-fitting and fiber count was discussed by Zhang and Laidlaw [10]; however, their technique did not address fascicle volume estimation and was limited to fascicles derived from deterministic algorithms.

BlueMatter is the first algorithm to produce a human brain projectome that combines diffusion-fitting and volume regularization. BlueMatter combines fascicle estimates from deterministic and stochastic DWI-FT algorithms; uniquely integrating algorithms with different strengths. Because BlueMatter has a very large and complex search space, the algorithm is highly parallelized and takes advantage of modern distributed computing architectures. Using this algorithm on a 2048-processor BlueGene/L supercomputer with 0.5 TB of memory we create the highest resolution, physically plausible projectome to date.

Table 1. Symbols and terms

Name	Description (Values)
C	Candidate fibers collected from many tractography algorithms
P, P_i	Current projectome estimate and portion of estimate intersecting ith voxel
E	Total projectome estimate error
E_1	Error between diffusion data prediction and measurement
E_2	Error from over filling voxels with white matter estimates
λ	Balance between E_1 and E_2 to compute E (**0.2**)
$A_{k,i}$	Diffusion attenuation measurement along the kth axis at the ith voxel
$\hat{A}_{k,i}$	Predicted diffusion attenuation along the kth axis at the ith voxel
$\nu_{C,i}, \nu_{U,i}$	Estimated volume of a voxel attributed to CSF and unoriented tissue
$\nu_{f,i}$	Estimated volume of a voxel attributed to the fascicle tissue
ν_T	Target volume of white matter
d_C, d_U	Diffusivity in all directions within canonical CSF and unoriented tissue compartments (**3.1, 0.85** $\mu m^2/ms$)
d_a, d_r	Diffusivity along the axial axis and radial axes within a canonical white matter fascicle compartment (**2.0, 0.275** $\mu m^2/ms$)
D_f	Canonical white matter diffusion tensor with eigenvalues d_a,d_r and d_r, and first eigenvector oriented with fascicle tangent
σ	Standard deviation of the data estimated across all brain voxels
τ_L	Convergence threshold for local improvements (**100 in 5000 iterations**)
τ_G	Convergence threshold for global improvements (**50 in 50 iterations**)
Δ_N	Amount of time to find their next neighbor state (**5 minutes**)

2 Algorithm

BlueMatter searches a massive collection of fascicle candidates C to select a projectome P. The quality of a projectome is evaluated by a global error function E. C is intended to be very large and contain a superset of the real fascicles. The collection C must be refined because it (a) contains many implausible fascicles, (b) reflects the biases of the generating tractography algorithms, and (c) is not optimized to fit the data or satisfy volumetric constraints. BlueMatter leverages the BlueGene/L architecture to search for an optimal projectome estimate from the enormous space of possible projectomes within C.

Table 1 defines key symbols and terms.

Error Terms. The BlueMatter algorithm penalizes a projectome, P, with a global error metric, E,

$$E = (1 - \lambda)E_1 + \lambda E_2. \tag{1}$$

The error E is a convex combination of E_1, which measures the difference between the predicted and observed diffusion weighted images, and E_2, the amount the volume is overfilled. The parameter λ modulates the balance between E_1 and E_2 and is selected empirically (Section 3). We seek solutions that predict the data to within the measurement noise; we also seek solutions with no more fascicles than a given voxel volume allows.

The projectome estimates the amount of volume in a voxel occupied by fascicles and the remaining space is then filled with isotropic diffusion according to the ratio $\nu_{U,i}/\nu_{C,i}$. The diffusion sensitization strength b is set by the MRI scanning sequence. The MRI sequence also specifies the axes of diffusion sensitivity or q_k, where $k \in [1, K]$.

We define the diffusion estimation error, E_1, and the local volume overfilling error, E_2, at voxel i as

$$\begin{aligned} E_{1,i} &= \tfrac{1}{K} \textstyle\sum_{k=1}^{K} \frac{(\hat{A}_{k,i} - A_{k,i})^2}{\sigma^2}, \\ E_{2,i} &= \begin{cases} 0 & \text{if } \sum_{f \in P_i} \nu_{f,i} < \nu_T \\ \sum_{f \in P_i} \nu_{f,i} - \nu_T & \text{otherwise.} \end{cases} \end{aligned} \tag{2}$$

Both E_1 and E_2 are the sums across voxels of $E_{1,i}$ and $E_{2,i}$, respectively. E_1 is normalized by the variance of the noise and thus represents error in units of data variance (or transformed into standard deviations). E_2 is in units of mm^3.

We compute E_1 based on a model for the predicted diffusion weighted image attenuation at voxel i and direction k, $\hat{A}_{k,i}$. BlueMatter uses the powder average of the separate compartments [11]. This term can be calculated from the projectome fibers passing through this voxel, P_i

$$\hat{A}_{k,i} = \frac{A_{0,i}}{\nu_{C,i} + \nu_{U,i} + \sum_{f \in P_i} \nu_{f,i}} (\nu_{C,i} \exp(-bd_C) +$$

$$\nu_{U,i} \exp(-bd_U) + \sum_{f \in P_i} \nu_{f,i} \exp(-bq_k^T D_{f,i} q_k)). \quad (3)$$

Parallel Candidate Generation. The set of candidate fibers is created by combining estimates from multiple tractography algorithms. It is intended that C contain a superset of fascicles; hence, we use a large candidate set derived from STT [3], TEND [12] and ConTrack [7] algorithms.

Projectome Search Overview. BlueMatter searches through C using a parallel stochastic hill climbing algorithm with multiple-restarts. A standard steepest ascent hill climbing algorithm would take the current state and every one of its neighbors; and choose the next state to be the neighbor that increases the inverse of the cost function the most [13]. By contrast, stochastic hill climbing algorithm only looks at a subset of the current state's neighbors and selects the next state that provides the largest improvement. For large neighborhoods and high data dimensionality, stochastic hill climbing substantially reduces the search time or in this case makes search feasible.

To reduce the likelihood of being stuck in local maxima, ridges or plateaus, groups of BlueGene/L processors are devoted to independent stocastic hill climbing processes (multiple-restarts) and the best resulting projectome is selected after each group converges. This follows the rule of thumb that when the terrain of the cost function is unknown it is beneficial to devote resources to covering more variable states rather than designing complex local movements. Thus, BlueMatter simultaneously exploits inherent data parallelism in the stochastic hill climbing algorithm and process parallelism in the multiple-restart approach.

Stochastic Search Implementation. For each restart, a subset of processors is selected to search the current projectome neighborhood. Each processor within a subset is assigned a random samples of C. The processors perform the following sequential optimization procedure for a specified amount of processor time.

To start, each processor selects a random voxel group. The processor then alternates between fiber addition and subtraction operations using the selected group. During addition, fibers that intersect the current voxel group are sampled from C. If the addition reduces the total error, the fiber is retained. During subtraction, fibers that intersect the voxel group are randomly sampled from P. If removing the fiber reduces the total error, the fiber is deleted. The algorithm alternates between addition and subtraction for a voxel group until the error reduction slows below a predefined threshold τ_L. After local improvement slows, a new voxel group is selected and the local optimization begins again.

Each processor optimizes its subset of candidate pathways, and processors within the subset exchange their current projectome estimate. After a small amount of processor time Δ_N has passed each processor points to a unique guess for the next projectome. The projectome with the lowest error is selected as the next state.

The volume constraint limits the size of the projectome state information, which is exchanged between processors quickly. Efficient communication is crucial to achieving the benefit of the massive and distributed database.

Multiple-Restart. The multiple-restart is achieved by devoting multiple sets of processors to independent runs of the stochastic hill climbing algorithm. When the rate of improvement of projectome error of each of the separate stochastic hill climbing groups slows below a threshold τ_G, the groups have converged to a projectome estimate. The projectome with the least error is selected as the optimal projectome.

The Initial Projectome. BlueMatter's initial projectome estimate is a subset of the STT and TEND fibers in the Mori human white matter atlas [14]; these are selected to minimize the error terms (see 2). This is a very small subset of candidate fibers (10,000) that are repeatedly found with previous algorithms.

3 Results

We first used a synthetic data set to demonstrate that BlueMatter recovers ground truth orientation and volume from diffusion-weighted images. We then used BlueMatter to estimate a projectome in human DWI data.

Synthetic. The synthetic dataset (Figure 1A) consists of two fascicle bundles. One bundle (blue) fills 60% of each voxel with fibers coherently organized in one direction. A second bundle (green) occupies 40% of each voxel with fascicles oriented perpendicular to the blue bundle. At the intersection, voxels are 100% filled with white matter from the two bundles. After accounting for the white matter, unfilled space within the bundles is filled with isotropic diffusion of the same mean diffusivity as white matter. The ends of both bundles are capped with voxels containing isotropic diffusion at the mean diffusivity of gray matter.

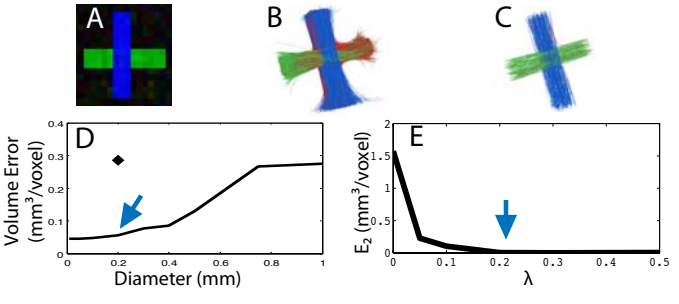


Fig. 1. Projectome estimation from synthetic data. (A) A synthetic data set comprising two fiber groups (green and blue). (B) A 3D view of the fascicle candidate database, C. (C) The BlueMatter projectome estimate P accurately recovers the synthetic fascicle ratio. (D) Small fascicle diameters produce a projectome with more accurate volume estimates. If probabilistic candidate pathways are excluded from C, the error is significantly worse (black diamond). (E) Increasing λ has a negligible effect on the diffusion-fitting error, E_1 (not shown) and significantly reduces E_2. Blue arrows (D and E) indicate the BlueMatter parameters used in the experiments.

The other voxels in the 20x20x20 volume are modeled as CSF. Diffusion data were simulated using the partial compartment model (Eq. 1) measured in six directions (plus a b=0 measurement) with 8 repeats. Rician noise with the same standard deviation as the data used in Section 3.2 (σ=48) was added to the magnitude attenuation signal.

Figure 1B shows the candidate set C derived using the methods described in Section 2. The fascicles are colored according to which ends they reach with blue and green matching the original; red fascicles are wrong turns. The candidate fascicles do not have the correct volume ratios and include incorrect fascicles. After running BlueMatter on this small synthetic data set, the proper blue-green ratio is returned and only a small number of wrong turns remain (Figure 1C).

We used this synthetic data to set (a) the fascicle diameter and (b) the balance between the two error terms. We evaluated the accuracy of the volume match by varying the modeled fascicle diameter (Figure 1D). A small fascicle diameter, on the order of 200μm (blue arrow), produces a good match to the synthetic projectome volume. Including probabilistic candidate pathways is important to achieving an accurate match. Without these candidate fascicles the volume estimate is significantly worse (black diamond).

Altering the balance between the diffusion-fitting and volume error terms has a negligible effect on the diffusion-fitting error (not shown). Setting the balance parameter, λ, reduces the volume error (E_2) error. We selected $\lambda = 0.2$ (blue arrow) for the experiments. With these parameters, BlueMatter recovers the volume and orientation of the synthetic white matter.

Human Projectome. In the second experiment we used BlueMatter to estimate a human projectome. First, 44,000 cortical and subcortical gray matter voxels were semi-automatically identified using FAST [15] and FIRST [16]. A set C containing 180 billion candidate fascicles was constructed. The set included fascicles from each and every gray matter voxel (Section 2). The BlueMatter search algorithm converged on a projectome with 200,000 fascicles after 9 days

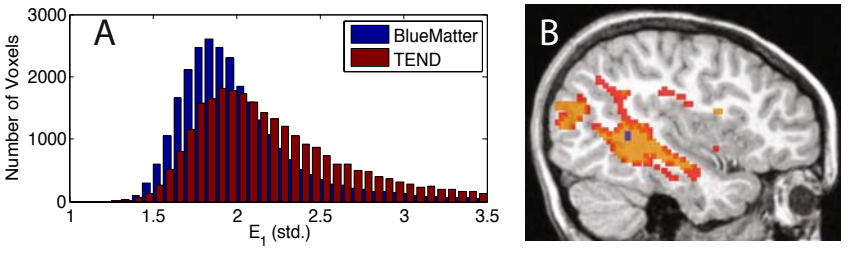


Fig. 2. Comparison between the BlueMatter and TEND human projectomes. (A) Diffusion-fit errors E_1 for BlueMatter (blue) and TEND (red) projectomes. BlueMatter reduces diffusion-fit error in most voxels. (B) A mid-temporal sagittal section of the human brain; the color overlay shows voxels overfilled by at least 2x (red) and 3x (orange) with the TEND projectome. The BlueMatter projectome only overfills two voxels (blue) by 2x and none by 3x.

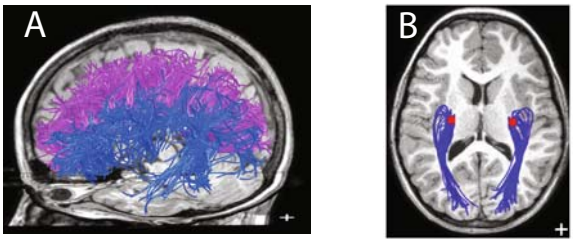


Fig. 3. Important fiber tracts within the BlueMatter projectome. (A) Pathways project from corpus callosum to lateral surface (blue). (B) Optic radiation (blue) connecting lateral geniculate nucleus (red) to primary visual cortex.

of compute time on the 2048-processor BlueGene/L supercomputer. We also computed a projectome using the TEND algorithm (see Izhikevich et al. [17]). The TEND projectome comprises 300,000 fascicles.

To estimate the white matter volume in each voxel of the TEND projectomes, we searched over fascicle model diameters (20-300μm) to find the diameter that has the lowest diffusion-fit error within the white matter core (linearity > 0.3). The optimal TEND diameter is the same as BlueMatter's diameter (200μm).

Using the optimal fascicle diameter, we could estimate the diffusion-fitting (E_1) and volume conservation (E_2) errors for both algorithms. The mean E_1 errors are 2.4 (mode 2.1) and 2 (mode 1.8) for the TEND and BlueMatter projectomes, respectively (Figure 2A). Using a scatter plot (not shown), we verified that BlueMatter E_1 error is better at nearly every voxel. The smallest possible mean E_1 error can be found by fitting a tensor to the data at each voxel: This lower bound is 1.7 (mode 1.6), which is slightly better than the BlueMatter fit.

We assessed the volume conservation error by counting the number of overfilled voxels (Figure 2B). The BlueMatter projectome has only 12 (200) voxels overfilled by more than 3x (2x); the TEND projectome has 3500 (8200) overfilled voxels. We explored compensating for TEND's overfilling by reducing the fascicle diameter. Reducing the diameter to 100μm reduces the overfilling error to match BlueMatter, but the mean E_1 value rises to 2.7 and many core white matter voxels, such as the corona radiata, are emptied of white matter.

Important Fiber Tracts. The BlueMatter human projectome contains the major intra-hemispheric white matter pathways as reported by Wakana et al. [14]. The projectome also contains fasciculi not present in the atlas a) fascicles that connect the corpus callosum to lateral cortex (Figure 3A, blue) and b) optic radiation (Figure 3B). These tracts are rarely found using local, greedy methods.

4 Conclusion

The BlueMatter algorithm searches across a pool of 180 billion candidate fascicles, drawn from many sources, to find a projectome. The highly parallelized implementation takes advantage of the 2048-processor BlueGene/L supercomputer.

BlueMatter is the first algorithm that simultaneously fits two global parameters of the projectome: the diffusion data (E_1) and volume conservation (E_2). The algorithm successfully identifies the optic radiation and other pathways that are frequently missed by local, greedy methods.

References

1. Braitenberg, V., Schüz, A.: Cortex: Statistics and Geometry of Neuronal Connectivity. Springer, Berlin (1998)
2. Kasthuri, N., Lichtman, J.W.: The rise of the 'projectome'. Nat. Methods 4(4), 307–308 (2007)
3. Basser, P.J., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A.: In vivo fiber tractography using DT-MRI data. Magn. Reson. Med. 44(4), 625–632 (2000)
4. Mori, S., van Zijl, P.C.: Fiber tracking: principles and strategies - a technical review. NMR Biomed. 15(7-8), 468–480 (2002)
5. Parker, G.J., Alexander, D.C.: Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. Philos Trans. R. Soc. Lond. B. Biol. Sci. 360(1457), 893–902 (2005)
6. Miller, N.R.: Diffusion tensor imaging of the visual sensory pathway: are we there yet? Am. J. Ophthalmol. 140(5), 896–897 (2005)
7. Sherbondy, A.J., Dougherty, R.F., Ben-Shachar, M., Napel, S., Wandell, B.A.: ConTrack: Finding the most likely pathways between brain regions using diffusion tractography. Journal of Vision 8(9), 1–16 (2008)
8. Poupon, C., Clark, C.A., Frouin, V., Regis, J., Bloch, I., Le Bihan, D., Mangin, J.: Regularization of diffusion-based direction maps for the tracking of brain white matter fascicles. Neuroimage 12(2), 184–195 (2000)
9. Peled, S., Friman, O., Jolesz, F., Westin, C.F.: Geometrically constrained two-tensor model for crossing tracts in DWI. Magn Reson Imaging 24(9), 1263–1270 (2006)
10. Zhang, S., Laidlaw, D.H.: Sampling DTI fibers in the human brain based on DWI forward modeling. In: Conf. Proc. IEEE Eng. Med. Biol. Soc., vol. 1, pp. 4885–4888 (2006)
11. Basser, P.J., Jones, D.K.: Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. NMR Biomed. 15(7-8), 456–467 (2002)
12. Lazar, M., Weinstein, D.M., Tsuruda, J.S., Hasan, K.M., Arfanakis, K., Meyerand, M.E., Badie, B., Rowley, H.A., Haughton, V., Field, A., Alexander, A.L.: White matter tractography using diffusion tensor deflection. Hum. Brain Mapp. 18(4), 306–321 (2003)
13. Russell, S., Norvig, P.: Artificial Intelligence: A Modern Approach, 2nd edn. Prentice-Hall, Englewood Cliffs (2003)
14. Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C., Mori, S.: Fiber tract-based atlas of human white matter anatomy. Radiology 230(1), 77–87 (2004)
15. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. IEEE Trans. Med. Imaging 20(1), 45–57 (2001)
16. Patenaude, B., Smith, S., Kennedy, M.: Improved surface models for FIRST. In: Proc. Hum. Brain Mapp. (2008)
17. Izhikevich, E.M., Edelman, G.M.: Large-scale model of mammalian thalamocortical systems. Proc. Natl. Acad. Sci. USA 105(9), 3593–3598 (2008)

Dual Tensor Atlas Generation Based on a Cohort of Coregistered non-HARDI Datasets

Matthan Caan[1,2], Caroline Sage[3], Maaike van der Graaf[1], Cornelis Grimbergen[1], Stefan Sunaert[3], Lucas van Vliet[2], and Frans Vos[1,2]

[1] Dept of Radiology, Academic Medical Center, University of Amsterdam, NL
[2] Quantitative Imaging Group, Delft University of Technology, NL
[3] Dept of Radiology, University Hospitals of the Catholic University of Leuven, BE
m.w.a.caan@tudelft.nl

Abstract. We propose a method to create a dual tensor atlas from multiple coregistered non-HARDI datasets. Increased angular resolution is ensured by random variations of subject positioning in the scanner and different local rotations applied during coregistration resulting in dispersed gradient directions. Simulations incorporating residual coregistration misalignments show that using 10 subjects should already double the angular resolution, even at a relatively low b-value of $b = 1000\,\mathrm{smm}^{-2}$. Commisural corpus callosum fibers reconstructed by our method closely approximated those found in a HARDI dataset.

1 Introduction

The white matter (WM) structure of the human brain can be studied by means of Diffusion Weighted Magnetic Resonance Imaging (DW-MRI). The diffusion is measured in multiple directions, from which the principal diffusion orientation field is reconstructed [1]. Fiber tracking throughout this field allows reconstruction of WM bundles, such as the corpus callosum [2].

The validity of local comparison of WM properties between different cohorts depends on good inter-subject anatomical correspondence. Variability in anatomical correspondence can be minimized by coregistration of the data on voxel basis. Both an affine and a non-rigid transformation can be computed to compensate for global and local variability, respectively [3].

Coregistration of data requires an appropriate reference space to be chosen. The reference space to which the subjects are transformed, can either be represented by one subject out of the cohort, or by an atlas that is built by averaging coregistered datasets. The ICBM FA atlas was generated by averaging 81 affinely transformed DTI datasets [4]. Recently, several DTI atlas-building approaches have been proposed with some yielding single subject-based atlases, e.g. [5], and others yielding population-based atlases [6].

By coregistering the data to such an atlas, fiber tracking can be performed in atlas space, which allows comparing tract statistics between patient and control groups [7]. However, tracking is known to be biased in regions where fibers are crossing. Fiber tracts may either deviate or truncate in such regions, due to

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 869–876, 2009.

inadequate modeling of the diffusion process by a single tensor. Fiber tracking through complex tissue regions is facilitated by higher order diffusion models, such as a dual tensor model or q-ball [8,9]. In order to apply such complex models, High Angular Resolution Diffusion Imaging (HARDI) data needs to be acquired [10]. One of the properties of a HARDI-acquisition is a high number of gradient directions, being more than 100. Clinical data is generally sampled at 6-45 gradient directions.

The design of a comparative study to find structural changes along fibers may be regarded as sampling a high-dimensional space, spanned by respectively three spatial coordinates, gradient directions and included subjects. Let us assume that the amount of available scanning time is constant, such as in a large population study in which only a few minutes of acquisition time is reserved for DW-MRI. In order to reduce bias in fiber tracking, one might prefer to acquire HARDI-data, yielding a denser sampling along the gradient axes. A fixed-time constraint implicitly imposes a lower sampling density along either the spatial or subject axes. Both are generally unacceptable, since not only small anatomical structures but also statistical power in the comparison need to be retained. Therefore, acquiring HARDI-data for clinical studies is not always realistic, as scanning time needs to be increased or scanner hardware needs to be upgraded to facilitate a higher SNR. Also, ongoing studies with conventional DW acquisition protocols, may not comply with the previously mentioned criteria.

In this study, we propose a method to fit a higher order diffusion model to a cohort of coregistered non-HARDI data. We consider the diffusion attenuation profiles of multiple subjects as realizations of one underlying fiber distribution. After non-rigid coregistration of the datasets, the resulting signal profiles are gathered to generate an artificial HARDI-dataset (disregarding the low b-value), with a high sampling density along the gradient axes. First, simulations incorporating residual coregistration misalignments are used to demonstrate the potential increase in angular resolution. Second, a dual tensor model is fitted to the data to generate a *dual tensor atlas*. We expect that in this atlas, tracking will be as accurate as in HARDI-data.

2 Method

2.1 Dual Tensor Model

We propose to estimate a dual tensor model from non-HARDI data of multiple subjects, with the purpose of precisely estimating two independent diffusion orientations per voxel. We assert a model for the diffusion weighted signal $S_j(\mathbf{q_j})$ of a subject j, with two tensors combined with an isotropic compartment:

$$\frac{S_j(\mathbf{q_j})}{S_{j,0}} = f_1 \exp\left(-\mathbf{q_j^T D_1 q_j}\right) + (1 - f_1 - f_{iso}) \exp\left(\mathbf{q_j^T D_2 q_j}\right) + f_{iso} \exp\left(-\mathbf{q_j^T q_j} D_{iso}\right). \tag{1}$$

Here $S_{j,0}$ is the signal for subject j measured without diffusion weighting and $f_{...}$ are the normalized volume fractions, while D_{iso} is the isotropic diffusion

constant. The vector $\mathbf{q_j} = \sqrt{b}\mathbf{g_j}$ reflects the diffusion weighting b in gradient direction $\mathbf{g_j}$. The diffusion tensors are spectrally decomposed, $\mathbf{D_{1,2}} = \mathbf{R_{1,2}ER_{1,2}}$, with $\mathbf{E} = \mathrm{diag}\left(\lambda_{\parallel}\ \lambda_{\perp}\ \lambda_{\perp}\right)$ being the eigenvalue matrix with axial and planar diffusion values (assuming axially symmetric tensors). The two rotation matrices $\mathbf{R_{1,2}}$ are parametrized using Euler angles α_{1-4}. The vector of eight parameter values $\boldsymbol{\theta} = \left(\lambda_{\parallel}\ \lambda_{\perp}\ \alpha_{1-4}\ f_1\ f_{iso}\right)$ is estimated in a least-squares sense.

2.2 Simulating Variation in the Cohort

A cohort of subjects that is studied for epidemiological processes or pathology is carefully defined by setting precise inclusion criteria. Thus, undesirable variations unrelated to the process of interest are minimized and may include handedness, sex, age and education [11]. Remaining variations are to be seen as normally distributed axial and planar diffusion values. Anatomical variations between subjects are apparent as differences in size and position of WM bundles. Non-rigid coregistration of the data accounts for the majority of these differences, residual misalignment of the principal diffusion direction was found to be 20° [12,13].

In an experiment on synthetic data, the expected angular resolution (i.e. the absolute error in the angle between the major axes of the tensors) will be computed, as a function of the number of included datasets. The abovementioned variations are incorporated by generating a signal profile per subject with normally perturbed eigenvalues and angles (using equation 1). Also, Rician noise is included to distort the simulated data. Gradient directions are randomly rotated per subject (see section 2.3). The parameter vector $\boldsymbol{\theta}$ is subsequently estimated on the combined data of multiple subjects. In this experiment, a diffusion weighting of $b = 1000\,\mathrm{smm}^{-2}$ is used to reflect our conventional clinical protocol.

2.3 Building a Dual Tensor Atlas

Coregistration. The DWI datasets of multiple subjects need to be transformed to a common reference space. In this study, a population-based atlas is build. To generate this atlas, all datasets are first affinely coregistered to the ICBM FA template [4]. As an intermediate step, we construct a (single tensor) DTI atlas that contains the full diffusion information [6]. The original DWIs were coregistered to our DTI atlas using an algorithm based on a viscous fluid model and optimization of mutual information as the coregistration criterion [12]. This coregistration consists of a combination of an affine and a non-rigid transformation to correct for global and local morphological differences, respectively. The effect of deformation discontinuities is removed by smoothing of the final deformation field with a Gaussian kernel of FWHM = 3mm.

Transformation. A diffusion-weighted image describes the MR signal attenuation due to diffusion in a specific gradient direction. If it is rotated, in our case during registration, signal correspondence with the gradient direction needs to be retained. This was done previously by fitting a single tensor to the data, after which preservation of principal direction (PPD) was applied to realign tensors

with the underlying fiber structure after transforming the tensor field [12]. The DWIs were then recomputed from the rotated tensor field. In the latter step, the number of degrees of freedom (DOF) is reduced to six (i.e. the DOF of the single tensor model), which prohibits a higher order model fit to the data. We therefore transform the original DWIs without intensity correction. Instead, the gradient directions are rotated, such that the vectors $\mathbf{q_j}$ become:

$$\mathbf{q_j} = \sqrt{b} \cdot \mathbf{R}_{\text{nonrigid,j}}(x, y, z) \cdot \mathbf{R}_{\text{affine,j}} \cdot \mathbf{g_j}. \tag{2}$$

The rotational component of the affine transformation $R_{\text{affine,j}}$ is applied globally, while the rotation due to non-rigid transformation $\mathbf{R}_{\text{nonrigid,j}}(x, y, z)$ is applied per voxel (x, y, z).

An advantage of our reorientation method is that both $R_{\text{affine,j}}$ and $\mathbf{R}_{\text{nonrigid,j}}$ will be dispersed among subjects (as demonstrated in figure 2(c)). Patient positioning in the scanner and inter-subject variations in bundle trajectories contribute to this distribution. The combined set of multiple subjects is sampled at a higher angular resolution than the original acquisition protocol. Hence, the merged dataset not only has an increased SNR, but indeed approaches a HARDI-dataset due to random dispersion of gradient directions per subject after coregistration.

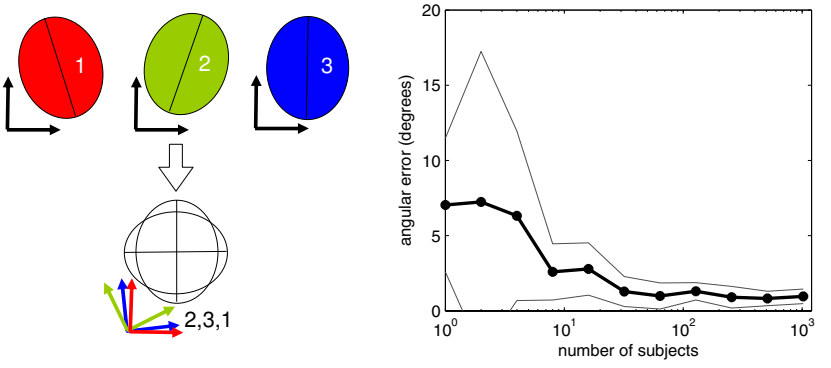
Tensor Estimation and Fiber Tracking. The dual tensor model can now be estimated (equation 1) on the merged dataset, using the dispersed sampling of gathered gradient directions. The least squares difference of the model and data is minimized using Levenberg-Marquardt optimization. Error-function parametrization ensures positiveness of the parameters. Next, WM fibers are reconstructed using High Angular Fiber Tracking (HAFT) [14]. As an extension to common streamline tracking algorithms, this method generates branches of fibers if within a voxel the angle between both tensors exceeds a certain threshold. One seed Region of Interest (ROI) and one AND-ROI through which fibers should pass have to be defined.

3 Results

3.1 Synthetic Data

Synthetic data was generated according to the model in equation 1. The eigenvalues were $\lambda_{\parallel} = 1.5 \cdot 10^{-3}\,\text{mm}^2\text{s}^{-1}$ and $\lambda_{\perp} = 0.4 \cdot 10^{-3}\,\text{mm}^2\text{s}^{-1}$ [15] and the angle between both tensors was 45°. For second and further realizations, the eigenvalues and the angle were normally perturbed with a standard deviation of 10%. S_0 was set to 230, whereas a volume fraction $f_1 = 0.5$ was used. Sixteen icosahedric and dodecahedric gradient directions on a half sphere [16] and along the 16 corresponding, antipodal directions were chosen, conform to our clinical scanning protocol. Rician noise (SNR = 20:1) distorted the data.

Next, data of different numbers of datasets with dispersed gradient directions were gathered and the dual tensor model was fit to the data. The isotropic volume fraction f_{iso} was included to account for inter-subject diffusion variations,



(a) (b)

Fig. 1. (a) Illustration of the dual tensor atlas generation, the gradient directions are randomly dispersed after coregistration, such that a dual tensor fit becomes possible. (b) Mean angular error (bold) with standard deviation (gray) in degrees as a function of the number of included datasets. The two tensors per subject were slightly perturbed, such that the angles were normally distributed with $\sigma = 5°$. The average angle between both tensors was 45°, 32 random gradient directions were used per subject.

whereas the isotropic diffusion was set constant: $D_{iso} = 0.5 \cdot 10^{-3}\,\mathrm{mm^2 s^{-1}}$. The mean and standard deviation of the angular error (defined in section 2.2) as a function of the number of subjects are depicted in figure 1. Initially, both the mean error and the standard deviation remained high when data of up to four subjects were used. For small sample sizes, modeled residual misalignments induce a slight fitting instability. Adding more subjects lowered the mean error and the standard deviation to $2° \pm 1°$ for 64 subjects. Further increase of data did not result in a lower error, again due to the applied perturbations.

This experiment shows that the angular resolution may be increased by estimating tensors based on a cohort.

3.2 Clinical Data

The proposed method was applied to a small cohort of 11 healthy subjects, of which DTI data were acquired on a 3.0T scanner (Intera, Philips, Best, The Netherlands). The spatial resolution was $2.0 \times 2.0 \times 2.2\,\mathrm{mm}$, such that 64 axial slices of matrix size 128x128 were acquired, while 32 gradient directions were used (see section 3.1) with a diffusion weighting of $b = 1000\,\mathrm{smm^{-2}}$. Additionally, one set of images was acquired without any diffusion-weighting. Eddy current distortions were corrected for by an affine registration in the phase direction [17]. For comparison, a HARDI-dataset was acquired, with 92 gradient directions and $b = 1800\,\mathrm{smm^{-2}}$. The acquired datasets were coregistered as described in section 2.3. The DWIs were resampled without intensity correction, and diffusion weighting vectors $\mathbf{q_j}$ were rotated (equation 2) and stored per voxel per deformation field of a subject. The merged dataset in total contained $32 \cdot 11 = 352$ DWIs.

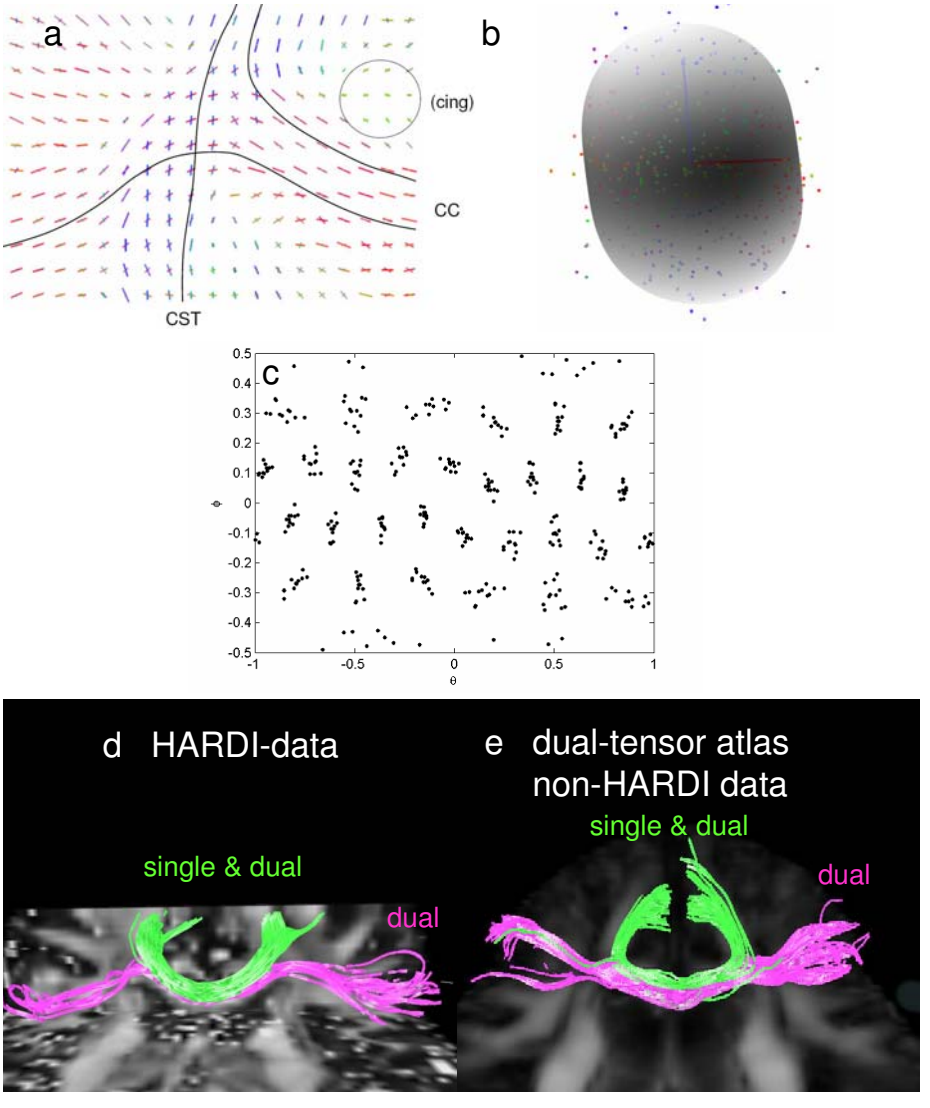


Fig. 2. Dual tensor atlas results. (a) Coronal cross-section through the estimated dual tensor atlas field, with principal diffusion orientations per voxel scaled by the volume fraction. Anatomical references are given (CST = corticospinal tract, CC = corpus callosum, cing = cingulum) together with an impression of bundle trajectories (drawn by the authors). (b) Aggregated ADC-profile plotted as points and fitted model in gray with principal orientations. Conventional color coding is used. (c) Aggregated gradient directions plotted in spherical coordinates. (d-e) HARDI-data and dual tensor atlas tracking result: part of the corpus callosum is tracked (green fibers), as well as commissural fibers (magenta fibers).

The dual tensor fit was initialized by positioning the principal eigenvectors in the plane of the first two eigenvalues of a single tensor fit. Dual tensor parameters were then estimated in each intercranial voxel. The results of the dual tensor atlas estimation are shown in figure 2. In figure 2(a), two distinct orientations within one voxel are mainly seen in the region where the corpus callosum is known to cross the corticospinal tract. Aggregated signal values within one voxel with the fitted model are depicted in figure 2(b). Due to the low b-value, low contrast is seen in the modeled signal. Moreover, significant noise and/or inter-subject variation may be observed (e.g. in the extension of the blue vertical vector in figure 2(b)). This confirmed the need for a sufficient amount of data in order to precisely estimate the orientation. In fact, high inter-voxel consistency in orientation in figure 2(a) is observed, indicating high fitting stability. Unfortunately, the gradient directions were not distributed uniformly, as displayed in figure 2(c). Still, a gain in angular resolution was achieved using our approach.

Fibers were reconstructed through the crossing and are displayed in figure 2(a). The fiber branching threshold was set to 12.5°, the angular threshold was 25° and the single tensor FA threshold 0.1. The obtained fibers in the HARDI-dataset and the dual tensor atlas are shown in figures 2(d) and (e) respectively. Both corpus callosum fibers (in green) as well as commissural fibers (in magenta) were successfully reconstructed in both datasets. A higher number of 38 commissural fibers was measured in the HARDI-dataset, compared to 11 in the dual tensor atlas. This is explained by the fact that in the HARDI-dataset a high number of fibers was running over the same trajectory.

4 Discussion

We proposed a method to create a dual tensor atlas from multiple coregistered non-HARDI datasets. Increased angular resolution was ensured by random variations of subject positioning in the scanner and different local rotations applied during coregistration resulting in dispersed gradient directions. Simulations showed that using 10 subjects should already double the angular resolution, even at a relatively low b-value of $b = 1000\,\mathrm{smm}^{-2}$. Experiments on clinical data revealed distinct orientations within a fiber crossing as well as inter-voxel consistency. Commisural corpus callosum fibers reconstructed by our method closely approximated those found in a HARDI dataset.

Although the adopted coregistration method uses DWIs generated with a single tensor model, as argued in section 2.3, we have apparently obtained a good correspondence in regions with crossing fibers. In a single tensor model, fiber crossings manifest themselves by an oblate spheroid. Although the oblate spheroid does not contain orientation information in the plane of the crossing, it permits accurate location of the crossing. The required orientational information for correct steering of the coregistration comes from its surrounding regions in which a single tensor model suffices.

We aim to apply our method to a large cohort in order to create an atlas that can be made available to the neuro-imaging community. WM bundles traveling through fiber crossings can then be included in comparative studies.

References

1. Basser, P., Mattiello, J., Bihan, D.L.: Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson. B (103), 247–254 (1994)
2. Wakana, S., Jiang, H., et al.: Fiber tract-based atlas of human white matter anatomy. Radiology 230, 77–87 (2004)
3. Maintz, J.B.A., Viergever, M.A.: A survey of medical image registration. Medical Image Analysis 2, 1–36 (1998)
4. Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A.V., Mahmood, A., Woods, R., Toga, A.W., Pike, G. B., Neto, P.R., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P., Mazziotta, J.: Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. Neuroimage 40, 570–582 (2008)
5. Ardekani, S., Sinha, U.: Statistical representation of mean diffusivity and fractional anisotropy brain maps of normal subjects. JMRI 24, 1243–1251 (2006)
6. van Hecke, W., Sijbers, J., D'Agostino, E., Maes, F., Backer, S.D., Vandervliet, E., Parizel, P., Leemans, A.: On the construction of an inter-subject diffusion tensor magnetic resonance atlas of the healthy human brain. NeuroImage 43, 69–80 (2008)
7. Goodlett, C., Fletcher, P., Gilmore, J., Gerig, G.: Group analysis of DTI fiber tract statistics with application to neurodevelopment. NeuroImage 45, S133–S142 (2009)
8. Descoteaux, M., Deriche, R., Knosche, T.R., Anwander, A.: Deterministic and probabilistic tractography based on complex fibre orientation distributions. IEEE Trans. Med. Im. 28, 269–286 (2009)
9. Kreher, B.W., Schneider, J.F., others: Multitensor approach for analysis and tracking of complex fiber configurations. Magn. Res. Med. 54, 1216–1225 (2005)
10. Tuch, D., Weisskoff, R., Belliveau, J.W., Wedeen, V.J.: High angular resolution diffusion imaging of the human brain. In: Proc 8th ISMRM, Philadelphia, vol. 321 (1999)
11. Caan, M., Vermeer, K., van Vliet, L., Majoie, C., Peters, B., den Heeten, G., Vos, F.: Shaving diffusion tensor images in discriminant analysis: A study into schizophrenia. Medical Image Analysis 10, 841–849 (2006)
12. Van Hecke, W., Leemans, A., d'Agostino, E., De Backer, S., Vandervliet, E., Parizel, P., Sijbers, J.: Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information. IEEE Trans. Med. Im. 26(11), 1598–1612 (2007)
13. Cao, Y., Miller, M., Mori, S., Winslow, R.L., Younes, L.: Diffeomorphic matching of diffusion tensor images. In: Proc. CVPR, pp. 67–76 (2006)
14. Toussaint, N., van Muiswinkel, A., Hoogenraad, F.G., Holthuizen, R., Sunaert, S.: Resolving fiber crossings: a two fiber model simulation result. In: Proc. ISMRM, vol. 13, p. 1339 (2005)
15. Pierpaoli, C., Jezzard, P., Basser, P.J., Barnett, A.: Diffusion tensor MR imaging of the human brain. Radiology 201, 637–648 (1996)
16. Akkerman, E.: Efficient measurement and calculation of MR diffusion anisotropy images using the platonic variance method. Magn. Reson. in Med. 49, 599–604 (2003)
17. Mangin, J.F., Poupon, C., Clark, C., Bihan, D.L., Bloch, I.: Eddy-current distortion correction and robust tensor estimation for MR diffusion imaging, pp. 186–193 (2006)

Estimating Orientation Distribution Functions with Probability Density Constraints and Spatial Regularity*

Alvina Goh[1], Christophe Lenglet[2], Paul M. Thompson[3], and René Vidal[1]

[1] CIS and Dept. of Biomedical Engineering, Johns Hopkins University
[2] CMRR and Dept. of Electrical and Computer Engineering, University of Minnesota
[3] LONI and Dept. of Neurology, University of California at Los Angeles

Abstract. High angular resolution diffusion imaging (HARDI) has become an important magnetic resonance technique for in vivo imaging. Current techniques for estimating the diffusion orientation distribution function (ODF), i.e., the probability density function of water diffusion along any direction, do not enforce the estimated ODF to be nonnegative or to sum up to one. Very often this leads to an estimated ODF which is not a proper probability density function. In addition, current methods do not enforce any spatial regularity of the data. In this paper, we propose an estimation method that naturally constrains the estimated ODF to be a proper probability density function and regularizes this estimate using spatial information. By making use of the spherical harmonic representation, we pose the ODF estimation problem as a convex optimization problem and propose a coordinate descent method that converges to the minimizer of the proposed cost function. We illustrate our approach with experiments on synthetic and real data.

1 Introduction

Diffusion magnetic resonance imaging (MRI) is a technique that produces in vivo images of biological tissues by exploiting the constrained diffusion properties of water molecules. An important area of research in diffusion MRI is the development of methods for reconstructing the orientation distribution function (ODF) – a *probability density function* (pdf) that characterizes the distribution of water diffusion along different directions on the sphere. A very successful reconstruction technique is high angular resolution diffusion imaging (HARDI) [1], which measures water diffusion along N uniformly distributed directions on the sphere. Given these signals, several reconstruction techniques can be used to characterize diffusion. Higher-order tensors leverage the work done in diffusion tensor imaging (DTI) [2] by using higher-order polynomials to model diffusivity [3,4]. [5] fits the HARDI signals with a mixture of tensors model whose weights are specified by a probability function defined on the space of symmetric positive definite matrices. Another approach is to construct the ODF directly from HARDI signals. One of the earliest methods, known as Q-ball imaging (QBI), uses the Funk-Radon transform to estimate ODFs [6]. ODFs have also been approximated with different basis functions

* Work supported by startup funds from JHU, by grants NSF CAREER IIS-0447739, NIH R01 HD050735, NIH R01 EB007813, NIH P41 RR008079, NIH P30 NS057091, ONR N00014-05-10836 and ONR N00014-09-1-0084.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 877–885, 2009.

such as spherical harmonics [7,8,9,10,11] and the Poussin kernel [12]. Such methods are typically very fast because the ODF can still be computed analytically.

A first important limitation of existing QBI methods is that they can give large diffusion estimates outside the major fiber directions. [8] addresses this by assuming that there is a distribution of fiber orientations in each voxel and using a sharpening spherical deconvolution method to transform the diffusion ODF into a sharp fiber ODF (fODF). A second limitation of existing QBI methods is that they do not enforce the estimated ODF to be nonnegative. When the diffusion MR signal is corrupted by noise, this can cause the estimated ODF to have negative values, a situation that does not obey the underlying principle of diffusion. [13] attempts to alleviate this problem by using a constrained spherical deconvolution (CSD) method to estimate the fODF. Even though CSD reduces the occurrence of negative values, it does not completely eliminate them. A more recent method [14] eliminates the negative values by minimizing a nonnegative least-squares cost function. A third limitation of existing QBI methods is that the ODF at each voxel is estimated independently of the information provided in the spatial neighborhood. This results in noisy estimates of the ODF field. While regularization methods have been developed [15,16], we are not aware of any work addressing all three issues for HARDI.

We present an estimation method that gives sharp diffusion ODFs, constrains the estimated ODF to be a proper pdf, and incorporates spatial regularization. Our algorithm is based on the ODF reconstruction scheme in [11], which derives the ODF taking into account the solid angle consideration and is able to give naturally sharp ODFs. This is different from existing works [8,10], where the computed ODF is the linear projection of the actual diffusion probability and gives an artificial weight to points according to their distances from the origin. Our method represents the ODF as a linear combination of spherical harmonic (SH) functions, whose coefficients are found by minimizing an energy that incorporates a regularization term and nonnegativity constraints. This results in a convex optimization problem whose global minimizer can be found using coordinate descent. We illustrate our method with experiments on synthetic and real data.

2 Analytical Computation of ODFs with Spherical Harmonics

We first review the ODF reconstruction scheme in [11]. Let S_0 be the baseline signal and $S(\theta,\phi)$ be the HARDI signal acquired at the gradient direction (θ,ϕ). The ODF is $\boldsymbol{p}(\theta,\phi) = \frac{1}{4\pi} + \frac{1}{16\pi^2} FRT\{\nabla_b^2 \ln(-\ln(\frac{S(\theta,\phi)}{S_0}))\}$, where FRT is the Funk-Radon transform and ∇_b^2 is the Laplace-Beltrami operator independent of the radial component. Notice that the first term integrates to 1 over the sphere, and the second term integrates to 0 [11]. The (modified) SH basis [8] of order l contains $R = \frac{(l+1)(l+2)}{2}$ terms defined for $j(k,m) = \frac{k^2+k+2}{2} + m$, $k = 0, 2, 4, \ldots, l$ and $m = -k, \ldots, 0, \ldots, k$, as

$$Y_j = \sqrt{2}\,\mathrm{Re}(Y_k^{|m|}) \text{ if } -k \le m < 0; \quad Y_k^0 \text{ if } m = 0; \quad \sqrt{2}\,\mathrm{Im}(Y_k^m) \text{ if } 0 < m \le k;$$

where $Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta)\mathrm{e}^{im\phi}$, $\theta \in [0,\pi], \phi \in [0,2\pi]$, P_l^m is a Legendre polynomial, and $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ are the real and imaginary parts, respectively. Notice that $Y_1(\theta,\phi) = \frac{1}{2\sqrt{\pi}}$ is a constant function on the sphere that integrates to a constant, whereas the integral of $Y_j(\theta,\phi), j > 1$ over the sphere is always 0.

In order to estimate the ODF, let $S(\theta_i, \phi_i)$ be the HARDI signal acquired at each of the N gradient directions, $(\theta_i, \phi_i)_{i=1}^N$, and define the $N \times 1$ vector $\boldsymbol{s} = \left[\ln(-\ln(\frac{S(\theta_1,\phi_1)}{S_0})) \ldots \ln(-\ln(\frac{S(\theta_N,\phi_N)}{S_0})) \right]^\top$. The signal $\boldsymbol{s}$ is first approximated as $\boldsymbol{s} \approx \mathbf{B}\mathbf{c}$, where $\mathbf{B}$ is the $N \times R$ SH basis matrix whose i-th row of $\mathbf{B}$ is given as $\mathbf{B}_i = \left[Y_1(\theta_i, \phi_i) \ \ldots \ Y_R(\theta_i, \phi_i) \right]$, and $\mathbf{c}$ is the $R \times 1$ vector of SH coefficients that parametrize the signal $\boldsymbol{s}$. Given $\boldsymbol{s}$ and B, the unknown vector $\mathbf{c}$ is found by solving the least-squares problem

$$\min_{\mathbf{c} \in \mathbb{R}^R} f(\mathbf{c}) = \frac{1}{2} \|\mathbf{B}\mathbf{c} - \boldsymbol{s}\|^2. \tag{1}$$

Assume now that the ODF is reconstructed using a tessellation scheme with M gradient directions, $(\theta_i^r, \phi_i^r)_{i=1}^M$. It is common for N, the number of gradient directions with which the HARDI signal is acquired, to be less than M. The reconstructed ODF is

$$\boldsymbol{p} = \mathbf{C}\mathbf{d}, \tag{2}$$

where $\mathbf{C}$ is an $M \times R$ SH basis matrix whose i-th row is $\mathbf{C}_i = \left[Y_1(\theta_i^r, \phi_i^r) \ldots Y_R(\theta_i^r, \phi_i^r) \right]$, and $\mathbf{d}$ is the vector of SH coefficients of the ODF, which is given by [11]

$$\mathbf{d} = \left[\frac{1}{2\sqrt{\pi}} \quad \mathbf{0}_{1 \times (R-1)} \right]^\top + \frac{1}{16\pi^2} \mathbf{L}\mathbf{P}\mathbf{c}. \tag{3}$$

$\mathbf{L}$ is the $R \times R$ diagonal Laplace-Beltrami eigenvalues matrix with $\mathbf{L}_{jj} = -l_j(l_j + 1)$, where l_j is the order of the j-th term, and $\mathbf{P}$ is the $R \times R$ diagonal Funk-Radon transform matrix, where $\mathbf{P}_{jj} = 2\pi P_{l_j}(0)$ and $P_{l_j}(0)$ is the Legendre polynomial of degree l_j at 0.

3 Nonnegative and Spatially Regularized ODF Estimation

Notice that while the least-squares estimation method in §2 enforces the sum of $\boldsymbol{p}$ to be one, it does not restrict $\boldsymbol{p}$ to be nonnegative. In addition, the ODF reconstruction at a voxel is done independently of the information contained in the spatial neighborhood of that voxel. In this section, we present our estimation method that constrains the estimated ODF to be a proper probability density function and incorporates spatial regularity.

Let V denote the HARDI volume and $|V|$ the number of voxels in V. At each voxel $\boldsymbol{x}_i = (x_i, y_i, z_i)$, we have the base-line signal $S_{0,i}$ and the $N \times 1$ HARDI signal S_i. Thus, we can define the signal vector $\boldsymbol{s}_i = \left[\ln(-\ln(\frac{S_i(\theta_1,\phi_1)}{S_{0,i}})) \ \ldots \ \ln(-\ln(\frac{S_i(\theta_N,\phi_N)}{S_{0,i}})) \right]^\top$ and its corresponding vector of SH coefficients $\mathbf{c}_i$. In order to enforce that the ODF $\boldsymbol{p}_i$ at $\boldsymbol{x}_i$ is nonnegative, we need to enforce the additional constraint $\boldsymbol{p}_i = \mathbf{C}\mathbf{d}_i \geq 0$. Making use of Eqns. (2) and (3), we rewrite the constraint as $-\mathbf{C}\mathbf{L}\mathbf{P}\mathbf{c}_i \leq 4\pi\mathbf{1}$.

To solve the ODF estimation problem in a way that accounts for the nonnegativity of $\boldsymbol{p}$ and incorporates spatial regularization, we define the following optimization problem

$$\min_{\mathbf{c}_1, \ldots, \mathbf{c}_{|V|}} g(\mathbf{c}_1, \ldots, \mathbf{c}_{|V|}) = \frac{1}{2} \sum_{i=1}^{V} \|\mathbf{B}\mathbf{c}_i - \boldsymbol{s}_i\|^2 + \lambda \sum_{\|\boldsymbol{x}_i - \boldsymbol{x}_j\| < r} w_{ij} \|\mathbf{c}_i - \mathbf{c}_j\|^2,$$
$$\text{subject to } -\mathbf{C}\mathbf{L}\mathbf{P}\mathbf{c}_i \leq 4\pi\mathbf{1}, \quad i = 1, \ldots, |V|, \tag{4}$$

The first term corresponds to the data term and the second term corresponds to the regularization term. The following parameters need to be set: 1) λ is the nonnegative regularization factor and marks the tradeoff between the data term and the regularization term.

When $\lambda \to 0$, spatial regularity is ignored whereas if $\lambda \to \infty$, the data term is disregarded. 2) r is the spatial radius defining the neighborhood of each voxel. Examples of defining the voxel connectivity in the 3D volume include choosing the 6 nearest voxels or those that lie within a certain distance. 3) w_{ij} are nonnegative weights that measure the similarity of the data at $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ within the local neighborhood. A common way of defining these weights is to use the Gaussian kernel and define $w_{ij} = \exp(-\frac{\|\boldsymbol{s}_i - \boldsymbol{s}_j\|^2}{\sigma^2})$.

It is possible to restate the optimization problem in Eqn. (4) as a large quadratic optimization problem and one could theoretically attempt to solve for $[\mathbf{c}_1, \ldots \mathbf{c}_{|V|}]$ simultaneously. However, adopting such a strategy will require a tremendous amount of memory and intensive computational power as a typical HARDI volume contains $|V| \approx 10^6$ voxels, the signal at each voxel is acquired at $N \approx 100$ gradient directions, and the ODF reconstruction is done with a few hundred tessellation directions. Instead, we adopt an iterative algorithm, specifically the coordinate descent method, and we show that coordinate descent will converge to the minimizer of Eqn. (4).

Theorem 1. *[17] Consider minimizing functions of the form*

$$\phi(\beta_1, \ldots, \beta_p) = \kappa(\beta_1, \ldots, \beta_p) + \sum_{k=1}^{p} \chi_k(\beta_k), \tag{5}$$

where β_k is a vector, $\kappa(\cdot)$ is a differentiable and convex function, and $\chi_k(\cdot)$ are convex functions. When the different vectors β_k's do not have overlapping entries and $\sum_{k=1}^{p} \chi_k(\beta_k)$ is separable, coordinate descent converges to the minimizer of $\phi(\cdot)$. The coordinate descent method is formally described as

1. *Initialization: Set $t = 0$ and choose any $\beta^0 = (\beta_1^0, \ldots, \beta_p^0) \in$ domain(ϕ).*
2. *At each iteration $t + 1, t \geq 1$: Given $\beta^t = (\beta_1^t, \ldots, \beta_p^t) \in$ domain(ϕ), choose an index $s \in \{1, \ldots, p\}$ and compute a new estimate $\beta^{t+1} = (\beta_1^{t+1}, \ldots, \beta_p^{t+1}) \in$ domain(ϕ) such that*

$$\beta_s^{t+1} = \arg\min_{\beta_s} \phi(\beta_1^t, \ldots, \beta_{s-1}^t, \beta_s, \beta_{s+1}^t, \ldots, \beta_p^t), \quad \textit{and} \quad \beta_k^{t+1} = \beta_k^t, \ \forall k \neq s.$$

Now, the optimization problem in Eqn. (4) is equivalent to minimizing the Lagrangian

$$\min_{\mathbf{c}_1, \ldots, \mathbf{c}_{|V|}} \phi(\mathbf{c}_1, \ldots, \mathbf{c}_{|V|}) = g(\mathbf{c}_1, \ldots, \mathbf{c}_{|V|}) - \sum_{i=1}^{|V|} \boldsymbol{\gamma}_i^\top (\mathbf{CLPc}_i - 4\pi\mathbf{1}), \tag{6}$$

where $\boldsymbol{\gamma}_i \geq 0$. Since $g(\mathbf{c}_1, \ldots \mathbf{c}_{|V|})$ is the sum of two quadratic functions, it is differentiable and convex. In addition, since any affine function is convex (and concave), $\boldsymbol{\gamma}_i^\top (\mathbf{CLPc}_i - 4\pi\mathbf{1})$ is convex. Finally, the different $\mathbf{c}_i$'s, belonging to different voxels, do not overlap with each other and $\sum_{i=1}^{|V|} \boldsymbol{\gamma}_i^\top (\mathbf{CLPc}_i - 4\pi\mathbf{1})$ is separable. From Theorem 1, it is immediate to see that coordinate descent will converge to the minimizer for Eqn. (6) or equivalently, the minimizer for Eqn. (4). In addition, at each iteration, $\mathbf{c}_i^{t+1} = \arg\min_{\mathbf{c}_i} \phi(\mathbf{c}_1^t, \ldots, \mathbf{c}_i, \ldots, \mathbf{c}_{|V|}^t) = \arg\min_{\mathbf{c}_i} g(\mathbf{c}_1^t, \ldots, \mathbf{c}_i, \ldots, \mathbf{c}_{|V|}^t) - \boldsymbol{\gamma}_i^\top (\mathbf{CLPc}_i - 4\pi\mathbf{1})$. Therefore, we can solve for $\mathbf{c}_i^{t+1}$ from the quadratic programming problem, $\mathbf{c}_i^{t+1} = \arg\min_{\mathbf{c}_i} g(\mathbf{c}_1^t, \ldots, \mathbf{c}_i, \ldots, \mathbf{c}_{|V|}^t) = \arg\min_{\mathbf{c}_i} \frac{1}{2}\|\mathbf{Bc}_i - \boldsymbol{s}_i\|^2 + \lambda \sum_{\|\boldsymbol{x}_i - \boldsymbol{x}_j\| < r} w_{ij} \|\mathbf{c}_i - \mathbf{c}_j^t\|^2$ subject to $-\mathbf{CLPc}_i \leq 4\pi\mathbf{1}$. Algorithm 1 gives our ODF estimation method in detail. Even though Algorithm 1 estimates one $\mathbf{c}_i$ in each iteration, it is possible to partition the problem to estimate a subset of $\mathbf{c}_i$'s simultaneously.

Algorithm 1. Nonnegative ODF Estimation with Spatial Regularization

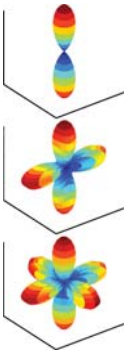
Given the HARDI volume $\boldsymbol{s}_i = [\ln(-\ln(\frac{S_i(\theta_1,\phi_1)}{S_{0,i}})), \ldots, \ln(-\ln(\frac{S_i(\theta_N,\phi_N)}{S_{0,i}}))]^\top, i = 1...|V|$, and predefined parameters: spatial radius r defining the voxel connectivity, threshold ϵ and maximum number of iterations P.

1. Calculate the nonnegative weights w_{ij}.
2. First pass $p = 0$ through the entire volume:
   At each voxel $\boldsymbol{x}_i$, calculate using a quadratic programming solver,
$$\mathbf{c}_i^0 = \arg\min_{\mathbf{c}} \frac{1}{2}\|\mathbf{Bc} - \boldsymbol{s}_i\|^2, \quad \text{subject to } -\mathbf{CLPc} \leq 4\pi\mathbf{1}.$$
3. Subsequent passes $p \geq 1$ through the entire volume:
   a. Initialize $t = 1$. While $t \leq |V|$,
      i. Set $i = t$ and compute using a quadratic programming solver a new estimate for voxel $\boldsymbol{x}_i$,
$$\mathbf{c}_i^{(p-1)|V|+t} = \arg\min_{\mathbf{c}} \frac{1}{2}\|\mathbf{Bc} - \boldsymbol{s}_i\|^2 + \lambda \sum_{\|\boldsymbol{x}_i - \boldsymbol{x}_j\| < r} w_{ij}\|\mathbf{c} - \mathbf{c}_j^{(p-1)|V|+(t-1)}\|^2,$$
$$\text{subject to } -\mathbf{CLPc} \leq 4\pi\mathbf{1}.$$
      ii. For the remaining voxels, set $\mathbf{c}_j^{(p-1)|V|+t} = \mathbf{c}_j^{(p-1)|V|+(t-1)}$ for $\forall j \neq i$.
      iii. Set $t \leftarrow t + 1$.
   b. Set $p \leftarrow p + 1$. The stop criterion used is when $p = P$ or the decrease in cost function $g(\cdot)$ between the p^{th} and $(p-1)^{th}$ iterations is less then ϵ.

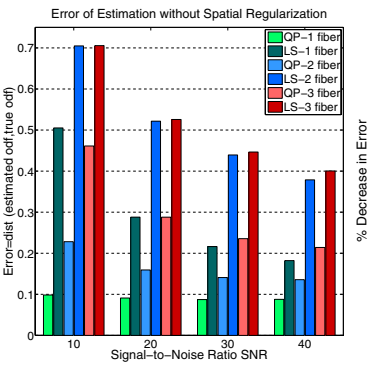
4 Experiments

We present experiments on synthetic and real datasets using the proposed estimation method. We examine the quality of the estimated ODF given by: 1) the commonly used least-square (LS) estimate obtained by solving Eqn. (1), 2) the estimate obtained with nonnegativity constraint solely (QP) by solving Eqn. (4) with $\lambda = 0$, and 3) the estimate obtained with nonnegativity constraint and spatial regularization (QP-S) by solving Eqn. (4) with non-zero λ. ODFs are computed using a $l = 6^{th}$ order SH expansion in all our experiments. Note that if the least-square solution results in an ODF that have negative values, the common treatment is to set such values to a small positive number.

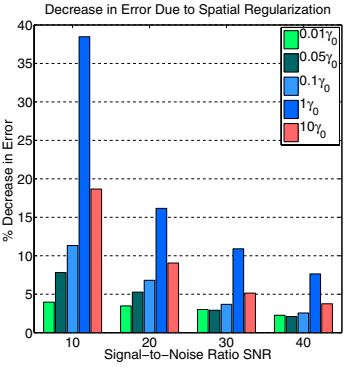
We first evaluate the performance on synthetic data generated using the multi-tensor method in [8]. The first synthetic experiment studies the improvement of the QP estimate over the LS estimate when the signal is corrupted by noise at a single voxel, in 100 trials. We first construct the true ODFs of $1, 2$, and 3 fibers, as shown in Fig. 1(a), and the corresponding HARDI signals $\{S(\theta_i, \phi_i)\}_{i=1}^N$ at $N = 100$ gradient directions with $S_0 = 1$. Noisy versions of $\{S(\theta_i, \phi_i)\}_{i=1}^N$ are generated by adding complex Gaussian noise with zero mean and standard deviation $\sigma = \frac{S_0}{\zeta}$, where ζ is the signal-to-noise ratio (SNR). We use the Riemannian distance $\text{dist}_{Rie}(\cdot,\cdot)$ between probability density functions [18] to compare different ODFs. Fig. 1(b) shows the error $\text{dist}_{Rie}(\boldsymbol{\phi}_t, \boldsymbol{\phi}_e)$ between the true ODF $\boldsymbol{\phi}_t$ and estimated ODF $\boldsymbol{\phi}_e$ using LS and QP when the voxel contains $1, 2$ or 3 fibers for varying levels of SNR. Notice that QP always gives a lower error than LS and does not give any negative values of $\boldsymbol{p}$ whereas LS results in 6%



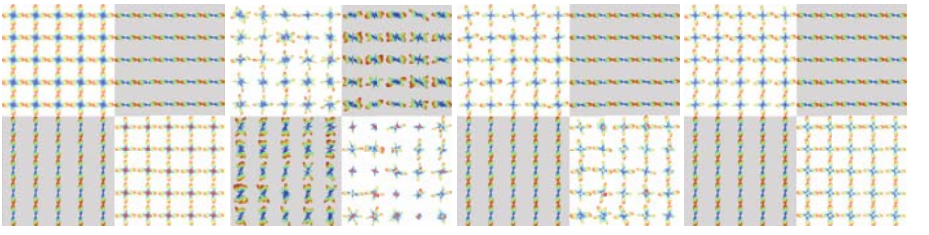
(a)



(b) Improvement of QP over LS



(c) Improvement of QP-S over QP



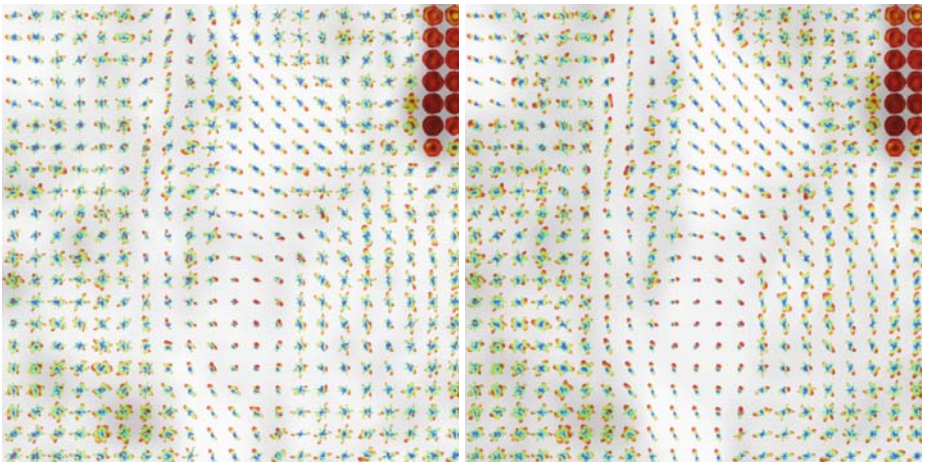
(d) True ODF field (e) LS estimate (f) QP estimate (g) QP-S estimate

Fig. 1. Synthetic experiments. The first experiment (Figs. 1(a)-1(b)) studies the improvement of the QP estimate over the LS estimate when the signal is corrupted by noise. The second experiment (Figs. 1(c)-1(g)) studies the effect of having spatial regularization under varying values of regularization parameter λ for the ODF field shown in Fig. 1(d).

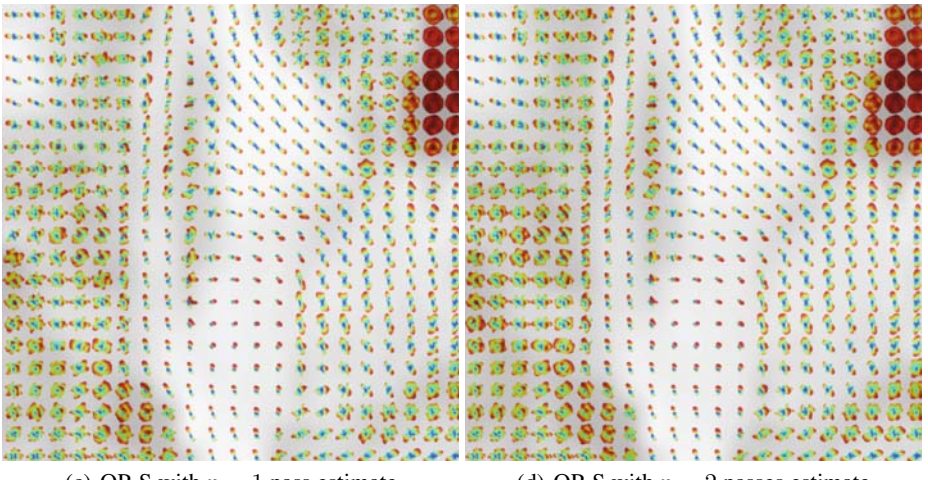
of the values being less than zero. In addition, the error increases when the number of fibers increases. This is expected as more acquisition directions are required to estimate an ODF with 3 fiber crossings compared to an ODF with a single fiber.

The second synthetic experiment shows the improvement of QP-S over QP in a ODF field over 100 trials. We construct an ODF field $\boldsymbol{\phi}_t$ as shown in Fig. 1(d). The voxels in the 1^{st} and 3^{rd} quadrants contain 1 fiber, the 2^{nt} quadrant 2 fibers and the 4^{th} quadrant 3 fibers (with 1 fiber pointing out of the plane). We study the effects of varying the regularization parameter λ from $0.01\lambda_0$ to $10\lambda_0$, where $\lambda_0 = 1$. Fig. 1(c) shows the $\%$ decrease in error when estimation is done with QP-S compared to only doing QP, for varying SNR and λ. The error is measured as $\sum_{\boldsymbol{x}} \text{dist}_{Rie}(\boldsymbol{\phi}_t(\boldsymbol{x}), \boldsymbol{\phi}_e(\boldsymbol{x}))$ where $\boldsymbol{\phi}_e$ is the estimated ODF field. At low SNR, QP-S gives an estimated ODF field that is significantly closer to the true ODF field. When λ is too small or too large, the resulting $\boldsymbol{\phi}_e$ is about the same quality of the LS estimate. Figs. 1(e)-1(g) show the estimated ODF fields for LS, QP, and QP-S with $p = 5$ passes of one trial at SNR$= 10$.

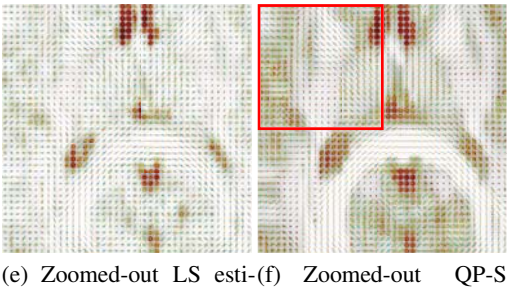
Finally, we apply our estimation method to a HARDI human brain dataset. Diffusion weighted MR images were obtained using the following imaging parameters: 55 axial slices (2mm thick), TR/TE=8250/92.3ms, with a 128×128 acquisition matrix (1.8mm in-plane resolution). 105 images were acquired, 11 with no diffusion



(a) LS estimate (b) QP estimate



(c) QP-S with $p = 1$ pass estimate (d) QP-S with $p = 2$ passes estimate



(e) Zoomed-out LS estimated ODF field (f) Zoomed-out QP-S with $p = 2$ estimated ODF field

Fig. 2. Estimation results for real brain data. Fig. 2(a) is the LS estimate, Fig. 2(a) the QP estimate, Figs. 2(b) and 2(c) are the QP-S estimates with $p = 1$ and $p = 2$ passes through the volume, respectively. Figs. 2(e)-2(f) show the zoomed-out LS ODF field and QP-S with $p = 2$ estimated ODF field with the results of the red box shown in Figs. 2(a)-2(d). The ODFs are superimposed on top of the generalized fractional anisotropy maps.

sensitization and $N = 94$ diffusion weighted images at $b = 1159$ s/mm^2. A portion of the results are shown in Fig. 2, where the ODFs are superimposed on top of the generalized fractional anisotropy maps. Figs. 2(a), 2(b), 2(c), and 2(d) show the estimates given by LS, QP, QP-S with $p = 1$ pass, and QP-S with $p = 2$ passes, respectively. Figs. 2(e) and 2(f) show the LS and QP-S with $p = 2$ estimated ODF fields of one slice where the zoomed-in results of the red box are shown in Figs. 2(a)-2(d). Notice that the ODF field estimated by LS is the noisiest, whereas QP-S with $p = 2$ gives a smoother ODF field and still preserves the discontinuities between different regions. In addition, the generalized fractional anisotropy map of QP-S shown in Fig. 2(f) is significantly cleaner than that of LS in Fig. 2(e) and the different regions of the brain can be seen more clearly.

5 Conclusion

We have presented an ODF estimation method that gives sharp diffusion ODFs, constrains the estimated ODF to be a proper pdf, and incorporates spatial regularization. Results on synthetic and real data demonstrate the advantage of working with our proposed algorithm. Future work will extend to multiple q-shell reconstruction method in [19].

References

1. Tuch, D.S.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magnetic Resonance in Medicine 48, 577–582 (2002)
2. Basser, P., Mattiello, J., LeBihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. Journal of Magnetic Resonance B 103, 247–254 (1994)
3. Barmpoutis, A., Jian, B., Vemuri, B.C., Shepherd, T.M.: Symmetric positive 4th order tensors & their estimation from diffusion weighted MRI. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 308–319. Springer, Heidelberg (2007)
4. Ghosh, A., Descoteaux, M., Deriche, R.: Riemannian framework for estimating symmetric positive definite 4th order diffusion tensors. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 858–865. Springer, Heidelberg (2008)
5. Leow, A.D., Zhu, S., Zhan, L., McMahon, K., de Zubicaray, G.I., Meredith, M., Wright, M.J., Toga, A.W., Thompson, P.M.: The tensor distribution function. MRM 61, 205–214 (2009)
6. Tuch, D.S.: Q-ball imaging. Magnetic Resonance in Medicine 52(6), 1358–1372 (2004)
7. Özarslan, E., Mareci, T.: Generalized diffusion tensor imaging and analytical relationships between DTI and high angular resolution diffusion imaging. MRM 50, 955–965 (2003)
8. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast and robust analytical Q-ball imaging. Magnetic Resonance in Medicine 58, 497–510 (2007)
9. Frank, L.R.: Characterization of anisotropy in high angular resolution diffusion-weighted MRI. Magnetic Resonance in Medicine 47(6), 1083–1099 (2002)
10. Hess, C.P., Mukherjee, P., Han, E.T., Xu, D., Vigneron, D.B.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. MRM 56(1) (2006)
11. Aganj, I., Lenglet, C., Sapiro, G.: ODF reconstruction in Q-ball imaging with solid angle consideration. In: IEEE Int. Symposium on Biomedical Imaging (2009)
12. Rathi, Y., Michailovich, O., Bouix, S., Shenton, M.: Orientation distribution estimation for Q-ball imaging. MMBIA (2008)

13. Tournier, J.D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion MRI. NeuroImage 35(4), 1459–1472 (2007)
14. Jian, B., Vemuri, B.: A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. IEEE TMI 26(11), 1464–1471 (2007)
15. Neji, R., Azzabou, N., Paragios, N., Fleury, G.: A convex semi-definite positive framework for DTI estimation and regularization. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Paragios, N., Tanveer, S.-M., Ju, T., Liu, Z., Coquillart, S., Cruz-Neira, C., Müller, T., Malzbender, T. (eds.) ISVC 2007, Part I. LNCS, vol. 4841, pp. 220–229. Springer, Heidelberg (2007)
16. Assemlal, H.E., Tschumperl'e, D., Brun, L.: Robust variational estimation of PDF functions from Diffusion MR signal. In: CDMRI (2008)
17. Tseng, P.: Coordinate ascent for maximizing nondifferentiable concave functions. Technical Report LIDS-P 1840, MIT, Laboratory for Information and Decision Systems (1988)
18. Goh, A., Lenglet, C., Thompson, P., Vidal, R.: A nonparametric Riemannian framework for processing high angular resolution diffusion images (HARDI). In: IEEE CVPR (2009)
19. Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N.: Multiple q-shell ODF reconstruction in q-ball imaging. In: Yang, G.-Z., et al. (eds.) MICCAI 2009, Part II. LNCS, vol. 5762, pp. 423–431. Springer, Heidelberg (2009)

Quantifying Brain Connectivity: A Comparative Tractography Study

Ting-Shuo Yo[1], Alfred Anwander[1], Maxime Descoteaux[2], Pierre Fillard[2], Cyril Poupon[2], and T.R. Knösche[1]

[1] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[2] Neurospin / CEA Saclay, Gif-sur-Yvette, France

Abstract. In this paper, we compare a representative selection of current state-of-the-art algorithms in diffusion-weighted magnetic resonance imaging (dwMRI) tractography, and propose a novel way to quantitatively define the connectivity between brain regions. As criterion for the comparison, we quantify the connectivity computed with the different methods. We provide initial results using diffusion tensor, spherical deconvolution, ball-and-stick model, and persistent angular structure (PAS) along with deterministic and probabilistic tractography algorithms on a human DWI dataset. The connectivity is presented for a representative selection of regions in the brain in matrices and connectograms.Our results show that fiber crossing models are able to reveal connections between more brain areas than the simple tensor model. Probabilistic approaches show in average more connected regions but lower connectivity values than deterministic methods.

1 Introduction

Diffusion-weighted magnetic resonance imaging (dwMRI) provides a non-invasive way to gain insight into the fibre architecture of the brain white matter, and thereby opens a window for the *in vivo* exploration of the anatomy of neural networks. In the past few years, a number of algorithmic approaches to the reconstruction of nerve fibre tracts from dwMRI have been proposed, collectively known as *tractography*. However, only few attempts have been made so far to quantitatively compare these different methods [1]. In this study, we compare a representative selection of state-of-the-art tractography algorithms, using *connectivity matrices* and *connectograms* based on a novel quantitative connectivity measure.

Most of the current techniques in dwMRI tractography can be divided into two major components: *local modeling* of the diffusion propagator or the fibre orientation structure in each voxel, and *fibre tracking algorithms* integrating this local information into streamlines representing fibre tracts.

Local modelling techniques convert the diffusion weighted MR signal into some quantity that can be used to determine the local fibre directions. There are two major classes of algorithms. The first one comprises methods aiming at a more or less simplified reconstruction of the diffusion propagator. Under the assumption of Gaussian anisotropic diffusion, this leads to the diffusion tensor (DT) model

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 886–893, 2009.

[2], which can represent only one main direction within each voxel and therefore fails to capture crossing or branchings of fibre populations. More complex models use e.g. compositions of ellipsoids or cylinders, like the multiple-tensor model [3] and the ball-and-stick model [4]. Another type of methods provides a less parameterized representation of the diffusion propagator. For example, if the q-space is completely sampled, one may use the spatial Fourier transform to reconstruct a restricted and blurred version of the diffusion propagator. This method is referred to as q-space imaging (QSI) or diffusion spectrum imaging (DSI) [5]. If only one b-value was used, one may compute the radial integral of the diffusion propagator (q-ball imaging, QBI, [6]) or its persistent angular structure (PAS, [7]).

The second class of methods directly aims at the reconstruction of the distribution of fibre orientations, e.g., by spherical deconvolution (SD) [8]. This approach requires an explicit model of the diffusion properties of a single fibre (convolution kernel). Its results are naturally more directly interpretable in terms of quantitative connectivity measures, as compared to methods that describe the diffusion propagator.

A detailed review of these methods can be found in [9]. In our comparison, we only include such local modelling methods that are suitable for high-angular resolution diffusion imaging (HARDI) data with a single b-value of b=1000. This naturally excludes QSI, which would require a complete Cartesian sampling of the q-space. Also QBI, which requires a higher b-value to provide a better angular discrimination than the tensor model [6], is not suitable for our dataset. The selection comprises the DT, multiple ball-and-stick, PAS and SD approaches. It can be considered representative, because it exemplifies all major approaches to the problem: (1) models assuming only one main fibre direction (DT), (2) models that allow for a small number of main fibre directions that has to be determined by some model selection procedure (multiple ball-and-stick), (3) models that represent the angular structure of the diffusion propagator (PAS), and (4) approaches that model the fibre orientation density directly, rather than the diffusion propagator (SD). Most other methods can be assigned to one of these classes (except QSI).

Based on these local models, tractography techniques integrate the local information connecting the voxels. There are two major approaches. With deterministic tractography, the reconstructed fibres are exclusively guided by the most likely directions in each voxel. In contrast, probabilistic fibre tracking methods repeat the streamlining process multiple times, each time with a new set of directions drawn from a probability distribution, which is based on the local model. In this study we evaluate each local model with both of these approaches, except for the multiple ball-and-stick model, which is only used with probabilistic tractography. The resulting collection of techniques also covers a range of software packages, which have been used in a large number of studies (see table 1).

For the quantitative comparison of the methods we focused on a type of information that is most naturally associated with tractography, namely if, and to what degree, two regions in the brain are connected by nerve fibres.This measure might also be a useful way to express prior information on connectivity

within various techniques for modelling functional networks in the brain, such as dynamic causal modelling [10].

The remainder of the paper is organized as follows. In section 2, the evaluated dwMRI tractography algorithms and the operational definition of the quantitative connectivity is presented in detail. The results of the experiments are shown in section 3, and details of difference among methods are discussed in section 4.

2 Methods

Dataset and Regions of Interests. All compared methods are applied to the HARDI dataset of one human subject. Diffusion images were acquired on a Siemens 3T Trio scanner with isotropic resolution of 1.7 mm (60 directions, b=1000s/mm2, GRAPPA/2, NEX3). Data is corrected for subject motion and registered to the anatomical T1 weighted image.

Fourteen language-related brain regions are selected as the regions of interests (ROI) for the quantification of anatomical connectivity (see Fig. 1). Eleven of the selected areas are located on the cortical sheet. In these cases the ROIs are placed at the interface between white and grey matter, which is defined as those voxels with fractional anisotropy (FA) greater than 0.15, which neighbour voxels with an FA of less than 0.15. Three additional regions comprise a mid-sagittal cross section of the corpus callosum, a horizontal cross section of the pyramidal tract, and the surface of the thalamus. The size of each selected ROI can also be found in figure 1.

Compared Algorithms. Table 1 summarizes all algorithms and software packages used for fibre reconstruction. The concept and implementation of each algorithm can be found in the corresponding references.

Definition of Quantitative Anatomical Connectivity. We define a measure, which reflects the influence the mean neuronal activity in one region has

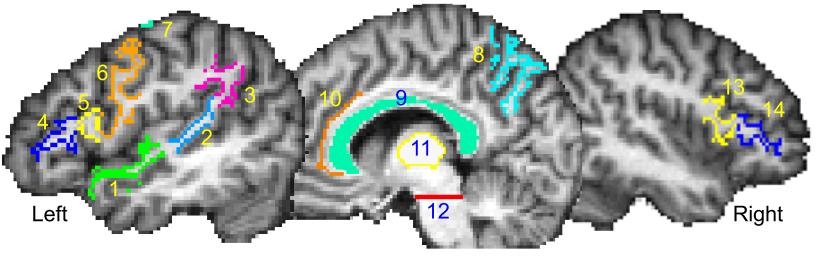


Fig. 1. Locations of selected regions of interests (ROIs). The names and sizes (number of voxels) are: 1. anterior superior temporal gyrus (STG) (497); 2. posterior STG (378); 3. angular gyrus (507); 4. Brodmann area 45 (BA 45) (319); 5. BA 44 (164); 6. precentral gyrus (PCG) ventral (796); 7. PCG dorsol (615); 8. precuneus (731); 9. corpus callosum (316);10. anterior cingulate (155);11. thalamus (385);12. cortical spinal tract (180);13. BA45, right hemisphere (400);14. BA44, right hemisphere (347).

Table 1. DWI tractography methods included in the comparison and related references. The computation time is mainly dominated by the local model fitting, and can be different due to the implementation.

Local Model	Comp. Time	Probabilistic Tractography	Deterministic Tractography
Diffusion tensor (DT [2])	∼10 sec.	[11]	MedINRIA[1] [12]
Multiple ball-and-stick [4]	∼2 days	FSL[2] [4]	-
Sph. deconv. (SD[8])	20-120 min.	MRtrix[3] [8]	BrainVisa[4] [13]
Per. ang. str. (PAS[7])	∼1 month	Camino[5] [14]	Camino [15]

on the mean activity in another region. If W_A and W_B are the sizes of the start and target regions, respectively (proportional to the number of output neurons as well as to the number of voxels), and F is the number of fibres connecting the two regions (proportional to the number of tracts, random walks, or similar), then the influence $C_{A\to B}$ of the mean activity of the start region N_A onto the mean activity of the target region N_B can be derived as follows. The mean activity N_B can be computed as the cumulative activity on the fibre tract N_F divided by the size of the target region W_B. The cumulative activity N_F is in turn proportional to the product of the mean activity of the start region N_A and the number of fibres F. The connectivity can then be computed as:

$$C_{A\to B} = \frac{N_B}{N_A} \propto \frac{F}{W_B}, \tag{1}$$

This connectivity measure is used throughout the comparison.

For probabilistic algorithms, we simulate fibres from each source point (e.g. voxel) n times. The connectivity is then computed as the ratio of the fibres that reach the target region divided by n times the number of source points in region B. For deterministic algorithms, we generate fibres starting from each voxel with FA$>$ 0.15 in the entire brain. We then count the number of fibres that run through A and B, as well as the number of fibres that run through A. The ratio of these two numbers multiplied by the ratio of the regions' sizes is then taken as an estimate of the connectivity.

3 Results

Figure 2 shows the logarithm of derived connectivity measures for each method in the matrix form. The ROIs are sorted with the spectral reordering algorithm [16] so that ROIs with high connectivity values will be clustered together. To

[1] http://www-sop.inria.fr/asclepios/software/MedINRIA/
[2] http://www.fmrib.ox.ac.uk/fsl/
[3] http://www.nitrc.org/projects/mrtrix/
[4] http://brainvisa.info/
[5] http://www.cs.ucl.ac.uk/research/medic/camino/

avoid confusion, all matrices are presented in the same ordering, which is based on the connectivity value derived from FSL. By comparing figure 2 with figure 1, we can see that anatomically closer areas are always clustered together, and the disconnection between the left and right hemisphere is very obvious.

Among different algorithms, the pattern of the connectivity matrix are similar, but the magnitude of connectivity values differ. A darkly shaded row of *cc*, which represents a high connectivity toward the corpus callosum, can be found throughout all methods. Also, the general pattern of highly connected regions is consistent across different methods.

The difference between deterministic (right column) and probabilistic (left column) tractography can also be seen in figure 2. Both deterministic tracking with DT and SD show significant white areas (i.e., no connection) in the matrices, while their probabilistic counterparts fill up almost the whole matrices. In addition, for SD, PAS and DT with probabilistic tractography, the shaded areas are lighter (i.e., lower connectivity) than those with deterministic fibre tracking algorithms.

Another way to visualize the quantitative connectivities is by graphs, called *connectograms*. The vertices are placed in the positions which approximately represent the locations of the ROIs, and the edges represent the magnitude of the connectivity measure. Figure 3 shows the resulting **connectograms**. In the connectograms, connectivity values above 10^{-1} are shown in red, 10^{-2} are in blue, and 10^{-3} in green. Edges with connectivity values bellow 10^{-3} are not shown, and the arrows represent the direction of connection. All methods show many arrows pointing toward vertex 9, corpus callosum, which is consistent with the heavily shaded rows in figure 2. The higher connectivity in deterministic tractography and more connections in more complicated local models can also be found in the connectograms. However, the PAS model shows less edges in probabilistic than deterministic tractography due to the in average lower connectivity values.

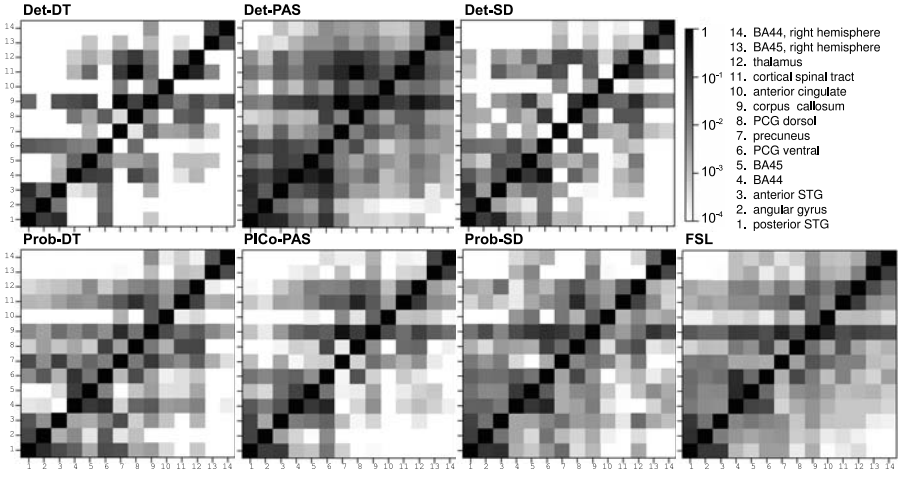


Fig. 2. Connectivity matrices derived from the collection of DWI tractography algorithms in logarithmic scale

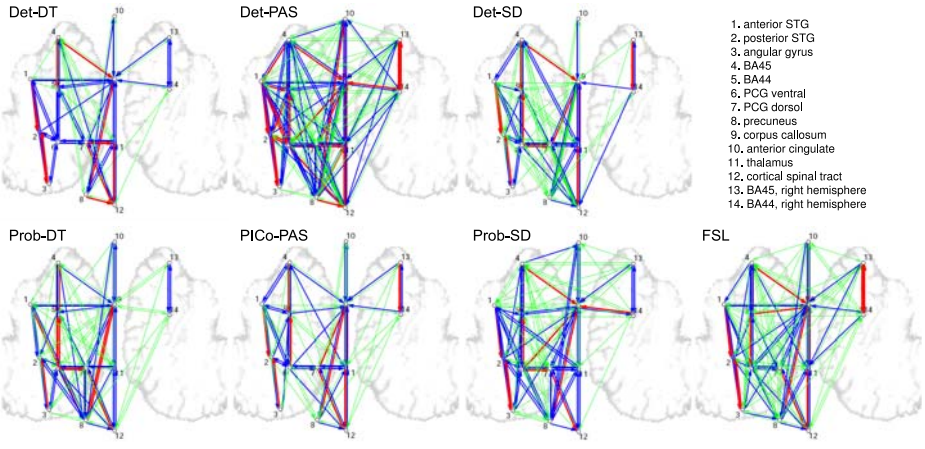


Fig. 3. Connectograms of DWI tractography algorithms. Connectivity values above 10^{-1} are shown in red, 10^{-2} are in blue, and 10^{-3} in green.

Due to the difference in theory and implementation, none of the compared methods give identical results to another. Nevertheless, there is clearly a great degree of similarity. The Mantel test is a technique used to estimate the resemblance between two proximity matrices computed about the same objects. This technique computes a covariant statistic between the two matrices, and then tests it against the null hypothesis of "no association" based on a non-parametric distribution obtained from permuting rows and columns together in one matrix. Several covariant statistics have been designed for different purposes, and we chose the Spearman rank correlation, ρ_M, as recommended in [17].

The Mantel test based on 1000 random permutations is applied to the connectivity matrices derived from all methods. As expected, all results are significantly correlated ($p = 0$, i.e., none of the 1000 permutations can produce a higher correlation). This result not only further confirms the similarity we observe from the shaded matrices and the connectograms, but also shows the proposed connectivity measure does retain certain structural information which is consistent across dwMRI tractography methods.

4 Discussion

Differences Among Methods. Considering the nature of the evaluated methods one would expect two major differences among the methods.

First, deterministic methods as compared to probabilistic ones are expected to feature a sparser connectivity matrix, i.e. there are fewer connections, but with higher connectivity values. This is due to the fact that probabilistic tractography produces a greater variability of fibre trajectories. In deterministic algorithms, fibres tend to follow the same trajectories to a much higher extent, resulting in more extreme connectivity values, i.e. two areas are more likely either strongly connected to not at all.

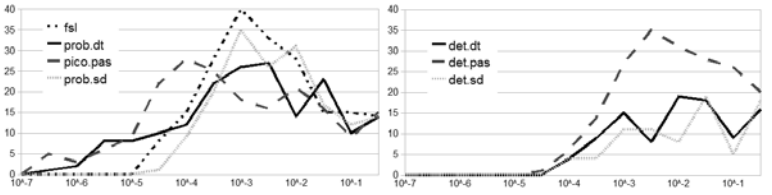


Fig. 4. Histograms of the connectivity values in logarithmic scale (zeroes are not counted). The left panel shows the histogram of the probabilistic tractography methods, and the right panel shows the deterministic ones.

Looking at the results, it turns out that this prediction clearly holds for the DT and SD model, but not for PAS. For PAS, the probabilistic PICo algorithm produces a clearly sparser connectivity matrix and connectogram than deterministic tracking. The reason for this could be the threshold we put at 10^{-4} in the matrix and 10^{-3} in the connectogram. Figure 4 shows the histograms of compared methods. It is clear that the histogram for PAS with deterministic tracking has a mode in 10^{-3}, and it with PICo in 10^{-4}.

The second prediction would be that, with the same tracking method, local models which can represent multiple fibre orientations (e.g. PAS, ball-and-stick, SD) will generate more connections between areas. This is logical because the additional fibre orientations might lead to new fibres that cross the major tracts. From our results, this prediction is only partially confirmed, since this trend is not as strong in the probabilistic tracking.

Concluding Remarks. In this study, we have compared a collection of state-of-the-art dwMRI tractography algorithms based on a quantitative connectivity measure. It has been shown that the proposed criterion give similar patterns across different methods, and also reasonably distinguish algorithms from each other. The results suggest that local models represent multiple fiber orientations can reconstruct more connections with a cost of more computation time (see table 1), as well as the probabilistic tractography. Since the difference in computational cost of tractography algorithms does not differ much, the choice of local models may dominate the computational resource required for this task. Although the optimal combination of methods can not be concluded from our findings, this study proposes a methodology to quatitatively compare different methods, which is of utmost importance for the community.

Future work will be focused on validating the comparison across different subjects, and to find a proper way to incorporate the quantified brain connectivity with other brain modelling techniques.

References

1. Descoteaux, M., Deriche, R., Knosche, T.R., Anwander, A.: Deterministic and probabilistic tractography based on complex fibre orientation distributions. IEEE Transactions on Medical Imaging 28(2), 269–286 (2009)

2. Basser, P.J., Mattiello, J., Lebihan, D.: Mr diffusion tensor spectroscopy and imaging. Biophys. J. 66(1), 259–267 (1994)
3. Parker, G.J., Alexander, D.C.: Probabilistic monte carlo based mapping of cerebral connections utilising whole-brain crossing fibre information. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 684–695. Springer, Heidelberg (2003)
4. Behrens, T.E., Berg, H.J., Jbabdi, S., Rushworth, M.F., Woolrich, M.W.: Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 34(1), 144–155 (2007)
5. Wedeen: Mapping fiber orientation spectra in cerebral white matter with fourier transform diffusion mri. In: Proc. of the 8th ISMRM, p. 82 (2000)
6. Tuch, D.S.: Q-ball imaging. Magn. Res. in. Med. 52(6), 1358–1372 (2004)
7. Jansons, K.M., Alexander, D.C.: Persistent angular structure: new insights from diffusion magnetic resonance imaging data. Inv. Prob. 19, 1031–1046 (2003)
8. Tournier, J.-D., Calamante, F., Connelly, A.: Robust determination of the fibre orientation distribution in diffusion mri: Non-negativity constrained super-resolved spherical deconvolution. NeuroImage 35(4), 1459–1472 (2007)
9. Alexander, D.: Multiple-fibre reconstruction algorithms for diffusion mri. Ann. N.Y. Acad. Sci. 1064, 113–133 (2005)
10. Friston, K.J., Harrison, L., Penny, W.D.: Dynamic Causal Modelling. NeuroImage 19(4), 1273–1302 (2003)
11. Anwander, A., Tittgemeyer, M., von Cramon, D., Friederici, A., Knosche, T.: Connectivity-based parcellation of broca's area. Cereb. Cor. 17, 816–825 (2007)
12. Toussaint, N., Souplet, J., Fillard, P.: Medinria: Medical image navigation and research tool by inria. In: Proc. of MICCAI 2007 Workshop, Australia (2007)
13. Cointepas, Y., Poupon, C., Maroy, R., Rivire, D., Le Bihan, D., Mangin, J.F.: A freely available anatomist/brainvisa package for analysis of diffusion mr images. In: Proc. 9th HBM CD-Rom Neuroimage, New York, vol. 19(2) (2003)
14. Parker, G.J., Haroon, H.A., Wheeler-Kingshott, C.A.: A framework for a streamline-based probabilistic index of connectivity (pico) using a structural interpretation of mri diffusion measurements. J. Mag. Res. Imag. 18(2), 242–254 (2003)
15. Cook, P.A., Bai, Y., Nedjati-Gilani, S., Seunarine, K.K., Hall, M.G., Parker, G.J., Alexander, D.C.: Camino: Open-source diffusion-mri reconstruction and processing. In: 14th Scientific Meeting of the ISMRM, p. 2759 (2006)
16. Higham, D.: Unravelling small world networks. Journal of Computational and Applied Mathematics 158(1), 61–74 (2003)
17. Schneider, J.W., Borlund, P.: Matrix comparison, part 2: Measuring the resemblance between proximity measures or ordination results by use of the mantel and procrustes statistics. J. Am. Soc.,Inf. Sci. and Technol. 58(11), 1596–1609 (2007)

Two-Tensor Tractography Using a Constrained Filter

James G. Malcolm[1], Martha E. Shenton[1,2], and Yogesh Rathi[1]

[1] Psychiatry Neuroimaging Laboratory, Brigham and Women's Hospital
Harvard Medical School, Boston, MA
[2] VA Boston Healthcare System, Brockton Division, Brockton, MA

Abstract. We describe a technique to simultaneously estimate a weighted, positive-definite multi-tensor fiber model and perform tractography. Existing techniques estimate the local fiber orientation at each voxel independently so there is no running knowledge of confidence in the estimated fiber model. We formulate fiber tracking as recursive estimation: at each step of tracing the fiber, the current estimate is guided by the previous. To do this we model the signal as a weighted mixture of Gaussian tensors and perform tractography within a filter framework. Starting from a seed point, each fiber is traced to its termination using an unscented Kalman filter to simultaneously fit the local model and propagate in the most consistent direction. Further, we modify the Kalman filter to enforce model constraints, *i.e*. positive eigenvalues and convex weights. Despite the presence of noise and uncertainty, this provides a causal estimate of the local structure at each point along the fiber. Synthetic experiments demonstrate that this approach significantly improves the angular resolution at crossings and branchings while consistently estimating the mixture weights. *In vivo* experiments confirm the ability to trace out fibers in areas known to contain such crossing and branching while providing inherent path regularization.

1 Introduction

The advent of diffusion weighted magnetic resonance imaging has provided the opportunity for non-invasive investigation of neural architecture. Using this imaging technique, neuroscientists can investigate how neurons originating from one region connect to other regions, or how well-defined these connections may be. For such studies, the quality of the results relies heavily on the chosen fiber representation and the method of reconstructing pathways.

To begin studying the microstructure of fibers, we need a model to interpret the diffusion weighted signal. Such models fall broadly into two categories: parametric and nonparametric. One of the simplest parametric models is the diffusion tensor which describes a Gaussian estimate of the diffusion orientation and strength at each voxel. While robust, this model can be inadequate in cases of mixed fiber presence or more complex orientations, and so to handle more complex diffusion patterns, various alternatives have been introduced: weighted mixtures [1,2,3,4], higher order tensors [5], and directional functions [6]. In contrast, nonparametric techniques estimate an orientation distribution function (ODF) describing an arbitrary configuration of fibers. For this estimation, several techniques have been proposed, among them Q-ball imaging

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 894–902, 2009.

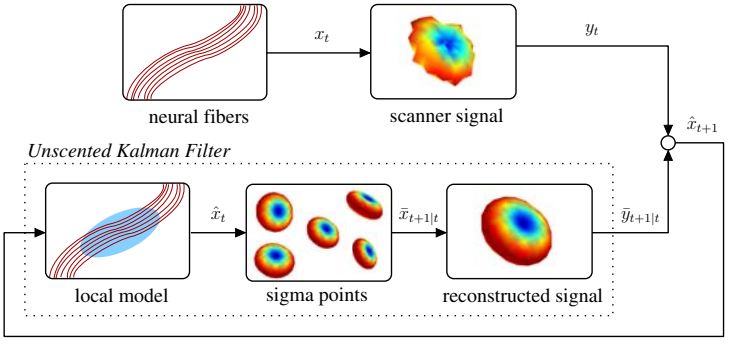


Fig. 1. System overview illustrating relation between the neural fibers, the scanner signal, and the unscented Kalman filter as it is used to estimate the local model. At each step, the filter uses its current model state to reconstruct a synthetic signal and then compares that against the actual signal from the scanner in order to update its internal model state.

[2], spherical harmonics [7,8], spherical deconvolution [9,10,11,6], and diffusion orientation transforms [12].

Based on these models, several techniques can be used to reconstruct pathways. Deterministic tractography using the single tensor model simply follows the principal diffusion direction, while multi-fiber models use various techniques for determining the number of fibers present or when pathways branch [3,13]. While parametric methods directly describe the principal diffusion directions, interpreting the ODFs from model independent representations typically involves a separate algorithm to determine the number and orientation of diffusion patterns present [14,9,8,15]. Several filtering approaches have been proposed. For example, Kalman and particle filters [16,17,18], as well as a moving least squares approach [19], have been used with single tensor streamline tractography, but these have been used for path regularization and not to estimate the underlying fiber model. One approach has used a linear Kalman filter, although this method was applied to estimate each voxel independently during acquisition [20].

1.1 Our Contributions

Of the approaches listed above, nearly all fit the model at each voxel independent of other voxels; however, tractography is a causal process: we arrive at each new position along the fiber based upon the diffusion found at the previous position. In this paper, we treat model estimation and tractography as such by placing this process within a causal filter. As we examine the signal at each new position, the filter recursively updates the underlying local model parameters, provides the variance of that estimate, and indicates the direction in which to propagate tractography.

To begin estimating within a finite dimensional filter, we model the diffusion signal using a weighted mixture of two tensors. This enables estimation directly from the raw signal without separate preprocessing or regularization. Because the signal reconstruction is nonlinear, we use the unscented Kalman filter to perform local model estimation and then propagate in the most consistent direction (Fig. 1). Further, we use a constrained version of the unscented Kalman filter to ensure the tensor eigenvalues are

positive and the mixture weights are non-negative and convex. Using causal estimation in this way yields inherent path regularization, consistent partial volume estimation, and accurate fiber resolution at crossing angles not found with independent optimization.

2 Approach

Section 2.1 provides the necessary background on modeling the measurement signal using tensors and defines the specific weighted two-fiber model employed in this study. Then, Section 2.2 describes how this model can be estimated using an unscented Kalman filter and further how the constraints are enforced.

2.1 Modeling Local Fiber Orientations

In diffusion weighted imaging, image contrast is related to the strength of water diffusion, and our goal is to accurately relate these signals to an underlying model of putative fibers. At each image voxel, diffusion is measured along a set of distinct gradients, $\mathbf{u}_1, ..., \mathbf{u}_m \in \mathbb{S}^2$ (on the unit sphere), producing the corresponding signal, $\mathbf{s} = [\, s_1, ..., s_m \,]^T \in \mathbb{R}^m$. For voxels containing a mixed diffusion pattern, a general weighted formulation is expressed as, $s_i = s_0 \sum_j w_j e^{-b\mathbf{u}_i^T D_j \mathbf{u}_i}$, where s_0 is the baseline signal intensity, b is an acquisition-specific constant, w_j are convex weights, and D_j are tensors, each representing a diffusion pattern.

From that general mixture model, we choose a restricted form with only two weighted components. This choice is guided by several previous studies which found two-component models to be sufficient at $b = 1000$ [2,3,13,14,4,21]. Also, we assume the shape of each tensor to be ellipsoidal, *i.e.* there is one dominant principal diffusion direction $\mathbf{m}$ with eigenvalue λ_1 and the remaining orthonormal directions have equal eigenvalues $\lambda_2 = \lambda_3$ (as in [4,6]). These assumptions leave us with the following model used in this study:

$$s_i = s_0 w_1 e^{-b\mathbf{u}_i^T D_1 \mathbf{u}_i} + s_0 w_2 e^{-b\mathbf{u}_i^T D_2 \mathbf{u}_i}, \tag{1}$$

where w_1, w_2 are convex weights and D_1, D_2 are each expressible as $D = \lambda_1 \mathbf{m}\mathbf{m}^T + \lambda_2 \left(\mathbf{p}\mathbf{p}^T + \mathbf{q}\mathbf{q}^T\right)$, with $\mathbf{m}, \mathbf{p}, \mathbf{q} \in \mathbb{S}^2$ forming an orthonormal basis aligned to the principal diffusion direction $\mathbf{m}$. The free model parameters are then $\mathbf{m}_1$, λ_{11}, λ_{21}, w_1, $\mathbf{m}_2$, λ_{12}, λ_{22}, and w_2. Lastly, we wish to constrain this model to have positive eigenvalues and convex weights ($w_1, w_2 \geq 0$ and $w_1 + w_2 = 1$).

2.2 Estimating the Fiber Model

Given the measured signal at a particular voxel, we want to estimate the underlying model parameters that explain this signal. As in streamline tractography, we treat the fiber as the trajectory of a particle which we trace out. At each step, we examine the measured signal at that position, use that measurement to update our model parameters within the filter, and propagate forward in the most consistent direction. Fig. 1 illustrates this filtering process.

To use a state-space filter for estimating the model parameters, we need the application-specific definition of four filter components:

Algorithm 1. Unscented Kalman Filter

1: Form weighted sigma points $\mathbf{X}_t = \{w_i, \mathbf{x}_i\}_{i=0}^{2n}$ around current mean $\mathbf{x}_t$ and covariance P_t with scaling factor ζ

$$\mathbf{x}_0 = \mathbf{x}_t \qquad \mathbf{x}_i = \mathbf{x}_t + [\sqrt{\zeta P_t}]_i \qquad \mathbf{x}_{i+n} = \mathbf{x}_t - [\sqrt{\zeta P_t}]_i$$

2: Predict the new sigma points and observations

$$\mathbf{X}_{t+1|t} = f[\mathbf{X}_t] \qquad \mathbf{Y}_{t+1|t} = h[\mathbf{X}_{t+1|t}]$$

3: Compute weighted means and covariances, *e.g.*

$$\bar{\mathbf{x}}_{t+1|t} = \sum_i w_i \, \mathbf{x}_i \qquad P_{xy} = \sum_i w_i (\mathbf{x}_i - \bar{\mathbf{x}}_{t+1|t})(\mathbf{y}_i - \bar{\mathbf{y}}_{t+1|t})^T$$

4: Update estimate using Kalman gain K and scanner measurement $\mathbf{y}_t$

$$\mathbf{x}_{t+1} = \bar{\mathbf{x}}_{t+1|t} + K(\mathbf{y}_t - \bar{\mathbf{y}}_{t+1|t}) \qquad P_{t+1} = P_{xx} - K P_{yy} K^T \qquad K = P_{xy} P_{yy}^{-1}$$

1. The system state ($\mathbf{x}$): the model parameters
2. The state transition function (f): how the model changes as we trace the fiber
3. The observation function (h): how the signal appears given a particular model state
4. The measurement ($\mathbf{y}$): the actual signal obtained from the scanner

For our state, we directly use the parameters for the two-tensor model in Eq. 1:

$$\mathbf{x} = [\, \mathbf{m}_1 \;\; \lambda_{11} \;\; \lambda_{21} \;\; w_1 \;\; \mathbf{m}_2 \;\; \lambda_{12} \;\; \lambda_{22} \;\; w_2 \,]^T, \quad \mathbf{m} \in \mathbb{S}^2, \lambda \in \mathbb{R}^+, w \in [0,1]. \quad (2)$$

For the state transition we assume identity dynamics; the local fiber configuration does not undergo drastic change as it moves from one location to the next. Our observation is the signal reconstruction, $\mathbf{y} = h[\mathbf{x}] = \mathbf{s} = [\, s_1, ..., s_m \,]^T$ using s_i described by the model in Eq. 1, and our measurement is the actual signal interpolated directly on the diffusion weighted images at the current position.

Since our signal reconstruction in Eq. 1 is nonlinear, we employ an unscented Kalman filter to perform estimation. Similar to classical linear Kalman filtering, the unscented version seeks to reconcile the predicted state of the system with the measured state and addresses the fact that these two processes–prediction and measurement–may be nonlinear or unknown. In Algorithm 1 we present the standard version of this filter; for more thorough treatments, see [22,23]. It is important to note that while particle filters are a common approach to nonlinear estimation, we chose instead the unscented Kalman filter primarily for its low computational complexity. With respect to state dimension, particle filters require the number of particles to be exponential to properly explore the state space. In contrast, the unscented filter requires $2n + 1$ particles (sigma points) for a Gaussian estimate of the n-dimensional state.

In this standard formulation, we have ignored the constraints on our model. This results in instabilities: the diffusion tensors may become degenerate with zero or negative eigenvalues, or the weights may become negative. To enforce appropriate constraints, one can directly project any unconstrained state $\mathbf{x}$ onto the constrained subspace [23]. In other words, we wish to find the state $\hat{\mathbf{x}}$ closest to the unconstrained state $\mathbf{x}$ which

still obeys the constraints, $A\hat{\mathbf{x}} \leq \mathbf{b}$. Using P_t as a weighting matrix, this becomes a quadratic programming problem:

$$\min_{\hat{\mathbf{x}}} (\mathbf{x} - \hat{\mathbf{x}})^T P_t^{-1} (\mathbf{x} - \hat{\mathbf{x}}) \quad \text{subject to} \quad A\hat{\mathbf{x}} \leq \mathbf{b}. \tag{3}$$

This projection procedure is applied within unscented Kalman filter procedure to correct at every place where we move in the state-space: after spreading the sigma points $\mathbf{X}_t$, after transforming the sigma points $\mathbf{X}_{t+1|t}$, and after the final estimate $\mathbf{x}_{t+1}$.

In this study, for voxels that can be modeled with only one tensor, we found it preferable to have both the tensor components similarly oriented. Upon encountering a region of dispersion, the second component is poised and ready to begin branching instead of having zero weight and arbitrary orientation. To favor such solutions, we require the weights of each of the components to be not just non-negative but also greater than 0.2, and so, in our current implementation, D and $\mathbf{b}$ are constructed to encode the following state constraints:

$$\lambda_{11}, \lambda_{21}, \lambda_{12}, \lambda_{22} > 0 \qquad w_1, w_2 \geq 0.2 \qquad w_1 + w_2 = 1. \tag{4}$$

3 Experiments

We first use experiments with synthetic data to validate our technique against ground truth. We confirm that our approach accurately recognizes crossing fibers over a broad range of angles and consistently estimates the partial volumes (Section 3.1). We then examine a real dataset to demonstrate how causal estimation is able to pick up fibers and branchings known to exist *in vivo* yet absent using other techniques (Section 3.2).

In these experiments, we compare against two alternative techniques. First, we use sharpened spherical harmonics with peak detection as described in [8] (order $l = 8$, regularization $L = 0.006$). This provides a comparison with an independently estimated nonparametric representation. Second, when performing tractography on real data, we also compare against single-tensor streamline tractography for a baseline.

3.1 Synthetic Validation

Following the experimental method of generating multi-compartment synthetic data found in [2,8,15], we averaged the eigenvalues of the 300 voxels with highest fractional anisotropy (FA) in our real data set: $\{1200, 100, 100\}\mu\text{m}^2/\text{msec}$. We used these eigenvalues to generate synthetic MR signals according to Eq. 1 at $b = 1000$ with 81 gradient directions on the hemisphere and introduced Rician noise (SNR $\approx$ 5 dB).

While the independent optimization techniques can be run on individually generated voxels, care must be taken in constructing reasonable scenarios to test the causal filter. For this purpose, we constructed a set of two-dimensional fields through which to navigate. In the middle is one long pathway where the filter starts at one end estimating a single tensor but then runs into voxels with two crossed fibers at a fixed angle and weighting. In this crossing region we calculated error statistics to compare against sharpened spherical harmonics.

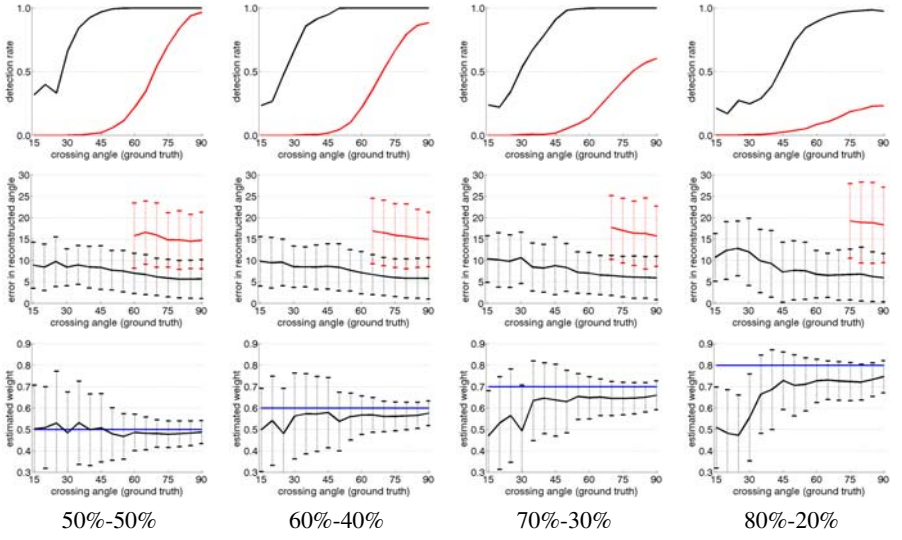
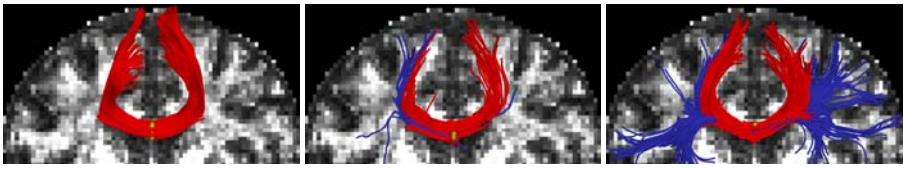


Fig. 2. Comparison of sharpened spherical harmonics *(red)* against filtered approach *(black)* over several different metrics: detection rate, angular resolution, estimated primary fiber weight *(rows, top to bottom)*. Each column is a different primary fiber weighting. The filter provides superior detection rates, accurate angular resolution, and consistent weight estimated. Trendlines indicate mean while dashed bars indicate one standard deviation.

From these synthetic sets, we examined detection rate, angular resolution, and estimated volume fractions and we plot the results in Fig. 2. Each column looks at a different primary-secondary weighting combination, and each row looks at a different metric. In the top row, we count how many times each technique distinguishes two separate fibers. The filtered approach *(black)* is able to detect two distinct fibers at crossing angles far below that using spherical harmonics *(red)*. Further, the filtered approach maintains such relatively high detection rates even at 80/20 partial voluming *(far right column)*. In the middle row, we look at where each technique reported two fibers and we record the error in estimated angles. From this, we see that spherical harmonics result in an angular error of roughly 15° at best and fails to detect a second component at angles below 60°. In contrast, the filtered approach has an error between 5-10° and is able to accurately estimate down to crossing angles of 30°. In the bottom row, we look at the primary fiber weight estimated by the filter. As expected, this estimate is most accurate closer to 90° *(blue line indicates true weight)*.

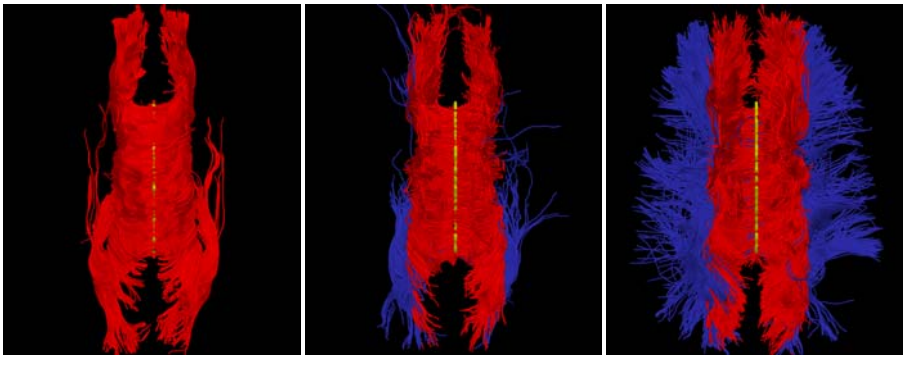
3.2 *In Vivo* Tractography

This study focuses on fibers originating in the corpus callosum. Specifically, we sought to trace out the lateral transcallosal fibers that run through the corpus callosum out to the lateral gyri. It is known that single-tensor streamline tractography only traces out the



Single-tensor Spherical harmonics Filtered

Fig. 3. Filtered tractography picks up many fiber paths consistent with the underlying structures. Both single-tensor streamline and sharpened spherical harmonics are unable to find the majority of these pathways. Seed region indicated in yellow.



Single-tensor Spherical harmonics Filtered

Fig. 4. Tracing fibers originating from the center of the entire corpus callosum viewed from above. The proposed filtered tractography is able to find many of the lateral projections *(blue)* while single-tensor is unable to find any and few are found with sharpened spherical harmonics. Seed region indicated in yellow.

dominant pathways forming the U-shaped callosal radiation while spherical harmonics only capture some of these pathways [8,15].

We begin by seeding each algorithm up to thirty times in voxels at the intersection of the mid-sagital plane and the corpus callosum. To explore branchings found using the proposed technique, we considered a component to be branching if it was separated from the primary component by less than 40° with FA$\geq$0.15 and weight above 0.3. Similarly, with sharpened spherical harmonics, we considered it a branch if we found additional maxima over the same range. We terminated fibers when either the generalized fractional anisotropy [2] of the estimated signal fell below 0.1 or the primary component FA fell below 0.15 or weight below 0.3.

We tested our approach on a human brain scan using a 3-Tesla magnet to collect 51 diffusion weighted images on the hemisphere at $b = 900\,\mathrm{s/mm}^2$, a scan sequence comparable those of [8,15]. Fig. 3 shows tracts originating from within a few voxels intersecting a chosen coronal slice. Confirming the results in [8,15], sharpened spherical harmonics only pick up a few fibers intersecting the U-shaped callosal radiata. In contrast, our proposed algorithm traces out many pathways consistent with the apparent

anatomy. Fig. 4 shows a view of the whole corpus callosum from above. The filtered approach is able to pick up many transcallosal fibers throughout the corpus callosum as well as infiltrating the frontal gyri to a greater degree than either alternate technique. To emphasize transcallosal tracts, we color as blue those fibers exiting a corridor of ± 22 mm around the mid-sagittal plane.

4 Conclusion

In this work, we demonstrated that using the unscented Kalman filter provides robust estimates of the fiber model with much higher accuracy than independent estimation techniques. Specifically, the proposed approach gives significantly lower angular error (5-10°) in regions with fiber crossings than using sharpened spherical harmonics (15-20°), and it reliably estimates the partial volume fractions.

References

1. Alexander, A., Hasan, K., Tsuruda, J., Parker, D.: Analysis of partial volume effects in diffusion-tensor MRI. Magnetic Resonance in Medicine 45, 770–780 (2001)
2. Tuch, D., Reese, T., Wiegell, M., Makris, N., Belliveau, J., Wedeen, V.: High angular resolution diffusion imaging reveals intravoxel white matter fiber heterogeneity. Magnetic Resonance in Medicine 48, 577–582 (2002)
3. Kreher, B., Schneider, J., Mader, I., Martin, E., Hennig, J., Il'yasov, K.: Multitensor approach for analysis and tracking of complex fiber configurations. Magnetic Resonance in Medicine 54, 1216–1225 (2005)
4. Peled, S., Friman, O., Jolesz, F., Westin, C.F.: Geometrically constrained two-tensor model for crossing tracts in DWI. Magnetic Resonance in Medicine 24(9), 1263–1270 (2006)
5. Basser, P., Pajevic, S.: Spectral decomposition of a 4^{th}-order covariance tensor: Applications to diffusion tensor MRI. Signal Processing 87, 220–236 (2007)
6. Kaden, E., Knøsche, T., Anwander, A.: Parametric spherical deconvolution: Inferring anatomical connectivity using diffusion MR imaging. NeuroImage 37, 474–488 (2007)
7. Anderson, A.: Measurement of fiber orientation distributions using high angular resolution diffusion imaging. Magnetic Resonance in Medicine 54(5), 1194–1206 (2005)
8. Descoteaux, M., Deriche, R., Anwander, A.: Deterministic and probabilistic Q-ball tractography: from diffusion to sharp fiber distributions. Technical Report 6273, INRIA (2007)
9. Jian, B., Vemuri, B.: A unified computational framework for deconvolution to reconstruct multiple fibers from diffusion weighted MRI. Trans. on Medical Imaging 26(11), 1464–1471 (2007)
10. Jansons, K., Alexander, D.: Persistent angular structure: New insights from diffusion MRI data. Inverse Problems 19, 1031–1046 (2003)
11. Tournier, J.D., Calamante, F., Gadian, D., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage 23, 1176–1185 (2004)
12. Özarslan, E., Shepherd, T., Vemuri, B., Blackband, S., Mareci, T.: Resolution of complex tissue microarchitecture using the diffusion orientation transform. NeuroImage 31(3) (2006)
13. Guo, W., Zeng, Q., Chen, Y., Liu, Y.: Using multiple tensor deflection to reconstruct white matter fiber traces with branching. In: Int. Symp. on Biomedical Imaging, pp. 69–72 (2006)

14. Zhan, W., Yang, Y.: How accurately can the diffusion profiles indicate multiple fiber orientations? A study on general fiber crossings in diffusion MRI. J. of Magnetic Resonance 183, 193–202 (2006)
15. Schultz, T., Seidel, H.: Estimating crossing fibers: A tensor decomposition approach. Trans. on Visualization and Computer Graphics 14(6), 1635–1642 (2008)
16. Gössl, C., Fahrmeir, L., Pütz, B., Auer, L., Auer, D.: Fiber tracking from DTI using linear state space models: Detectability of the pyramidal tract. NeuroImage 16, 378–388 (2002)
17. Björnemo, M., Brun, A., Kikinis, R., Westin, C.-F.: Regularized stochastic white matter tractography using diffusion tensor MRI. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 435–442. Springer, Heidelberg (2002)
18. Zhang, F., Goodlett, C., Hancock, E., Gerig, G.: Probabilistic fiber tracking using particle filtering. In: Medical Image Computing and Computer Assisted Intervention (MICCAI), pp. 144–152 (2007)
19. Zhukov, L., Barr, A.: Oriented tensor reconstruction: Tracing neural pathways from diffusion tensor MRI. Visualization, 387–394 (2002)
20. Poupon, C., Roche, A., Dubois, J., Mangin, J.F., Poupon, F.: Real-time MR diffusion tensor and Q-ball imaging using Kalman filtering. Medical Image Analysis 12(5), 527–534 (2008)
21. Behrens, T., Johansen-Berg, H., Jbabdi, S., Rushworth, M., Woolrich, M.: Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? NeuroImage 34, 144–155 (2007)
22. Julier, S., Uhlmann, J.: Unscented filtering and nonlinear estimation. IEEE 92(3), 401–422 (2004)
23. Simon, D., Simon, D.: Kalman filtering with inequality constraints for turbofan engine health estimation. IEE Proc.–Control Theory and Appl. 153(3), 371–378 (2006)

Characterization of Anatomic Fiber Bundles for Diffusion Tensor Image Analysis

Rubén Cárdenes, Daniel Argibay-Quiñones, Emma Muñoz-Moreno, and Marcos Martin-Fernandez*

Laboratory of Image Processing, University of Valladolid, Spain
{ruben,dargqui,emunmor,marcma}@lpi.tel.uva.es

Abstract. In this paper we deal with the problem of quantification of diffusion tensor (DT) data sets. A set of measures and a 2D tract mapping technique are proposed to analyze the fiber structures in brain white matter and to allow for comparisons between different subjects, either patients or controls. Features such as integrity, discontinuity and connectivity of the fiber bundles are proposed and analyzed, taking into account longitudinal and transverse information of the fiber bundle under study. The performance of the proposed characterization framework is shown analyzing the corticospinal tracts of control data sets and pathological cases, comparing the measures between controls and patients and also between the right and left hemispheres. A reproducibility study is also performed to show the robustness of the proposed measures.

1 Introduction and State of the Art

Diffusion tensor imaging (DTI) is a modality that measures the diffusion of water molecules in tissues, and allows to visualize the fiber structure of the brain, since water diffusion is constrained by myelin coat of the axons. Using the main diffusion direction at each voxel, the fiber trajectories can be estimated and then visualized in order to give to the specialist a better understanding of the white matter structure. This technique, known as fiber tracking or tractography, has been studied in many works before, and it can be used to compute measures along a given set of tracts to quantitatively analyze the connectivity in the brain.

DTI analyses have been performed to assess the effect of neurological diseases in the fiber structure. In most cases, scalar measures such as the fractional anisotropy (FA) or the mean diffusivity (MD) have been computed in the whole brain or in regions of interest. However, recent works propose to compute such measures in a given fiber tract of interest. To compare tracts between controls and patients, the same tract should be identified in every volume under study. Usually, seed regions are manually placed to generate these tracts. However, some atlas-based approaches [1,2,3,4] have been proposed to automatically identify the tracts of interest, and then measures are computed along such tracts. For

* This work has been funded by the national spanish grant TEC-2007-67073/TCM, supervised by the *Hospital Clínico Universitario de Valladolid* and the data sets have been acquired at *Centro de Diagnóstico Recoletas, Valladolid*.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 903–910, 2009.

instance, in [2] the authors compare tracts among healthy subjects and patients with temporal lobe epilepsy. The work in [4] measures the FA, MD and eigenvalues along the pyramidal tract. Measures along fibers are also proposed [5], such as fiber length, weighted fiber length, etc. to analyze the integrity of the white matter. Other interesting approach that avoids the use of fiber tracking is the *Tract Based Spatial Statistics* (TBSS) [6], but this method reduces the fiber structure of the brain to a projection onto the skeleton of the FA.

In this paper we propose a quantification method for anatomical fiber bundles using an automatic tractography method. We will show the application of this method in the pyramidal tract, although other fiber bundles can be used. We compare the obtained measures between a control group and a group of patients with neuro-degenerative diseases, as well as between the hemispheres of patients affected by a tumor with a healthy hemisphere.

2 Automated Extraction of Fiber Tracts

The fiber tract extraction is an atlas based method. For this reason we have constructed first a DTI model of the brain using 30 control healthy subjects. The data sets used have been acquired in a GE Signa 1.5 T MR scanner using 15 gradient directions, b = 1000 s/mm^2, 1.015x1.015x3 mm of voxel size, TR=9999.9 ms, TE = 80.90 ms, NEX=8, and spanning the entire brain. For the model construction we have registered all the Diffusion Weighted Images (DWI) of control subjects to the same reference framework. Then the DWI images were averaged, to finally estimate the tensor using the averaged DWI data. This model is then used to define a set of regions of interests (ROIs) each of them corresponding to one fiber bundle of anatomical interest. The ROIs are identified in the model using the white matter atlas in [7].

To automatically obtain the tracts of a given subject, the ROIs obtained for the model are deformed to fit the subject data. This deformation is a non linear transformation, that is obtained by means of a multi-resolution template matching registration between the model FA and the subject FA volumes.

Then a tractography method is executed using the voxels in the deformed ROIs as seeds. In this work a fourth order Runge-Kutta method is employed, due to its efficiency. In order to obtain more reliable tracts, a *brute force* approach is followed: the fiber tracts for the whole brain are computed, using a FA threshold of 0.15, and each fiber tract is assigned with an index value. Therefore each voxel with a FA value greater than 0.15 will point to an array of indexes corresponding to the fibers that pass through it. Then, the fibers selected for each ROI are those that pass through the voxels that belongs to that particular ROI.

Note that this methodology of fiber tracts identification does not warp the fibers in the subject, and this is done to perform robust measures over the original fibers computed from each subject, and not over the warped fibers (as in [3]), or over a set of voxels in the skeleton (as in [6]). Of course, this methodology is not error free, and is affected by the accuracy of the tractography and the registration. Although both methods can be improved, the effect of registration

errors has negative impact only in the case that the deformed regions do not cover completely the target regions in the subject data, because only the voxels with FA>0.15 are finally considered. Also, more complex methods can be used to solve fiber crossings, but it will affect drastically the execution time and the acquisition requirements, making this method non practical for clinical use.

3 Fiber Tracts Quantification

The goal of this work is to obtain quantitative measures that allows to discriminate between normal and abnormal fiber bundles, using the automatic tractography described above. It is important to highlight that the so computed fibers are not warped, and due to the automatic method to extract them the user variability is removed from the process.

3.1 Longitudinal Measures

We propose three new measures computed along the fiber paths that we will call longitudinal measures. As an example, we show in figure 1 (a) and (b) the average FA profiles obtained in the trajectory followed by two fiber bundles. They correspond to the right pyramidal tracts of a healthy subject (a) and a tumor patient (b). Notice the different behavior presented by these two profiles.

- **Tract integrity measure**. The first measure proposed is the integrity of the fiber tracts, and is defined using the FA, because it is directly related to the number of axons aligned in a predominant direction. Therefore, the FA is a good descriptor here, and the proposed integrity measure for a given tract is defined as the total amount of FA computed at the points involved in the fiber paths, divided by the number of fibers in the tract:

$$I = \frac{1}{M} \sum_{f=1}^{M} \sum_{i=1}^{N_f} \mathrm{FA}_f(x_i),$$

  where FA_f is the FA profile of fiber f, x_i are its fiber points, N_f is the number of points of fiber f, and M is the number of fibers computed. Notice that short fibers will contribute with low values to this measure which is desirable because we want to compare similar fiber bundles, and therefore we want to take into account length discrepancy between tracts. For this reason the mean FA is not used.
- **Tract discontinuity measure**. The integrity measure is related to the amount of directional diffusion in a fiber bundle. But if a fiber tract is affected in small zones along its path, its integrity could be slightly reduced by a negligible quantity. Thus, discontinuities or local changes along the fibers will have small impact on the previous measure. In order to consider this effect, we define the discontinuity of a fiber tract as the number of local minima of the fiber FA profiles, weighted by their distance to the maximum

FA value, which is 1. Individual fiber measures are averaged as before, to obtain a unique measure for the whole tract. If y_j are the positions of the local minima of each FA profile, in the fiber tract considered, the discontinuity measure is

$$D = \frac{1}{M} \sum_{f=1}^{M} \sum_{j=1}^{L_f} (1 - \mathrm{FA}_f(y_j)),$$

where again M is the number of fibers computed and L_f is the number of local minima obtained for fiber f. In order to avoid non relevant local minima, the noise of the FA profile is reduced with gaussian filtering. Low values of this measure means more continuous and reliable fibers.

– **Tract connectivity measure**. Finally, the measures obtained before are used to define a connectivity measure, that is associated with the fiber tract and with two ROIs. For the pyramidal tract case, two different ROIs are defined (one of them is placed at the internal capsule, and the other one at an axial plane above the corpus callosum). For the aforementioned reasons, each fiber tract will have different features, and just counting the number of them connecting the two ROIs will give rise to erroneous information. For instance for two different subjects, the same number of fibers connecting the two defined ROIs can be very similar, but the integrity and the discontinuity of the fibers can be quite different between them. For this reason the connectivity measure is based on the two measures proposed before:

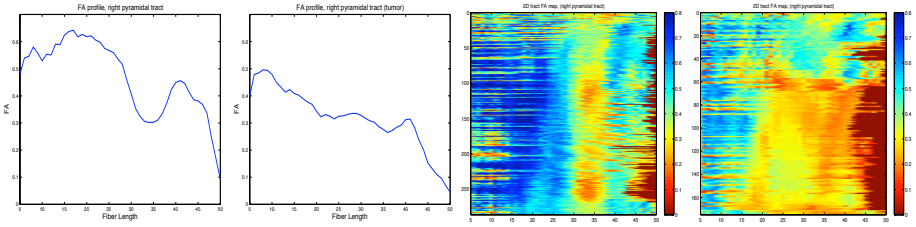
$$C = \frac{M \cdot I}{1 + D}$$

which increases with the integrity, and decreases with the discontinuity.

3.2 2D Tract Mapping

The defined measures are longitudinal and do not use the transverse direction or information between neighbor fibers. Hence, we propose to introduce information about this transverse direction using a 2D tract mapping method to visualize and measure the FA profiles in a two dimensional way. This consists in projecting the fiber tracts in a plane. Starting from the seeds, the FA along the fibers are stored in a 2D matrix where the rows represent the longitudinal dimension and the columns represent the transverse dimension. The implementation of this mapping is easy in the pyramidal tract, because the fibers can be ordered from anterior to posterior direction, and the longitudinal dimension is ordered from inferior to superior. Figure 1 (c) and (d), shows an example of these maps for the pyramidal tract of a control and a patient with a tumor.

We can therefore use this mapping to obtain again the measures proposed before, but instead of using only the longitudinal dimension, we can also use the transverse dimension to have more reliable measure results, using the map as a 2D image. The new measures are then defined as:



(a) (b) (c) (d)

Fig. 1. Average FA profiles in the right pyramidal tract of a control (a) and a patient with a tumor in this hemisphere (b), and corresponding 2D FA tract maps (c) and (d)

$$I_{2d} = \frac{1}{M} \sum_{i=1}^{N} \frac{\mathrm{FA}(\boldsymbol{x_i})}{1 + \alpha_l \left| \frac{\partial \mathrm{FA}(\boldsymbol{x_i})}{\partial x} \right| + \alpha_t \left| \frac{\partial \mathrm{FA}(\boldsymbol{x_i})}{\partial y} \right|}$$

where now FA is a 2D image, $\boldsymbol{x_i}$ are its pixels, N is the total number of pixels in the 2D map [1], α_l and α_t are weights used for the longitudinal and transverse dimensions respectively. In this work we will use $\alpha_l = 1$ and $\alpha_t = 0.5$ to give more relevance to the longitudinal dimension which is a more reliable dimension.

$$D_{2d} = \frac{1}{M} \sum_{j=1}^{L} |\nabla \mathrm{FA}(\boldsymbol{y_j})|(1 - \mathrm{FA}(\boldsymbol{y_j}))$$

where $\boldsymbol{y_j}$ are the pixels that corresponds to the same local minima as in the longitudinal discontinuity measure, and L is the total number of them. Finally, the connectivity 2D measure is:

$$C_{2d} = \frac{M \cdot I_{2d}}{1 + D_{2d}}$$

In order to explain better the measures behavior, 4 synthetic semicircular tracts, shown in figure 2, are analyzed. Comparison between the C_{2d} values obtained: 1079, 1040, 885 and 423, against the average FA values: 0.915, 0.830, 0.904, 0.918, from a to d respectively, demonstrate that C_{2d} provides a more natural ordering and a higher sensitivity to discriminate between different tracts.

4 Experiments

In this section several experiments are presented to show the performance of our quantification method. The used data has been acquired with the same parameters described in section 2.

[1] Due to difference in fiber sizes, N is greater than the number of points in the fiber tract, but there is no contribution of these additional pixels because they are zero.

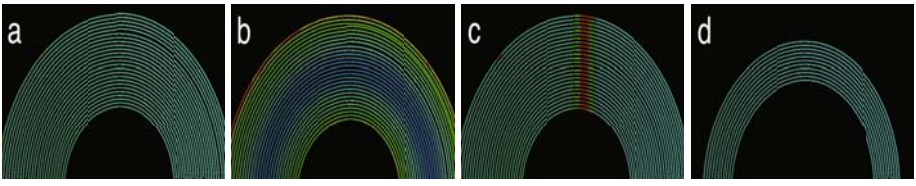


Fig. 2. Synthetic tracts used to explain the measures behavior

4.1 Real Case Study

The measures here proposed have been computed in the pyramidal tract of a group of 10 control subjects and in a group of 7 patients affected by a neurodegenerative disease (3 multiple sclerosis patients and 4 epilepsy patients). The mean and standard deviation from both groups are shown in table 1, and represented in figure 3. It is clear from these results that the integrity values are lower in the patient group, and the discontinuity is higher in the patient group as well. Consequently the connectivity measures obtained are also higher in the control group. The difference of the mean values between controls and patients are statistically significant in the integrity and connectivity measures with a significance level of $p < 0.05$, using a t-test. This is not the case for the mean FA or the number of fibers. For the left pyramidal tract, for instance, the mean FA values are 0.48 ± 0.02 and 0.45 ± 0.04 for the control and patients group respectively, and there are not statistical significant differences between both groups. With respect to the number of fibers, we obtain 222 ± 39 for controls and even higher values in the patient group: 242 ± 57, therefore it is not a valid measure for comparisons.

Other interesting study has been made to compare the measures in the two hemispheres of the same patient. In this case two patients affected with a tumor in only one hemisphere are considered. We show in figure 4 the pyramidal tracts of a control and a patient, using a colormap for the FA between 0.2 and 0.7. It is clear that although the fiber structure is similar in both cases, the FA is highly reduced in the patient due to the tumor. The connectivity values are also shown in figure 4 for the set of 10 controls used in the previous study and for the two patients. Values for the left pyramidal tract are in general higher than

Table 1. Quantitative measures. Mean values (and standard deviation) in the control and patients group.

Measure (side)	I (R)	I (L)	D (R)	D (L)	C (R)	C (L)
Control	21.53 (1.4)	22.53 (1.5)	0.37 (0.21)	0.23 (0.23)	3322 (697)	4126 (955)
Patient	18.74 (1.2)	19.34 (1.0)	0.50 (0.18)	0.40 (0.35)	2733 (823)	3653 (1671)
Measure (side)	I_{2d} (R)	I_{2d} (L)	D_{2d} (R)	D_{2d} (L)	C_{2d} (R)	C_{2d} (L)
Control	18.83 (1.2)	19.74 (1.3)	0.42 (0.21)	0.31 (0.17)	2808 (584)	3412 (825)
Patient	16.34 (1.0)	16.88 (0.8)	0.59 (0.16)	0.48 (0.42)	2238 (617)	3061 (1445)

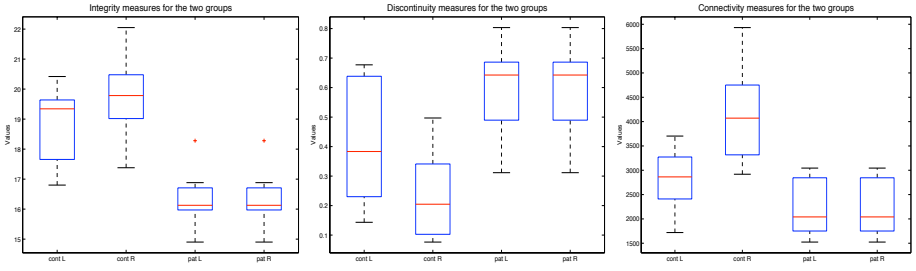


Fig. 3. Measures for the left and right pyramidal tracts of the two groups studied. Integrity 2D, Discontinuity 2D and Connectivity 2D.

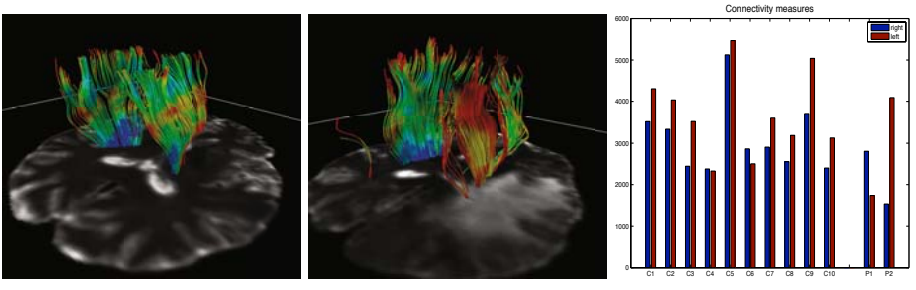


Fig. 4. Differences between the pyramidal tract in two hemispheres. Right: control subject; middle: tumor patient; left: connectivity measures in 10 controls and 2 patients.

Table 2. Mean and std values in 3 different DTI data sets from the same volunteer

Measure(side)	I (R)	I (L)	D (R)	D (L)	C (R)	C (L)
Mean (std)	22.92 (0.45)	23.61 (0.34)	0.21 (0.15)	0.13 (0.04)	3729 (198)	4627 (179)
Measure(side)	I_{2d} (R)	I_{2d} (L)	D_{2d} (R)	D_{2d} (L)	C_{2d} (R)	C_{2d} (L)
Mean (std)	19.89 (0.49)	20.56 (0.22)	0.27 (0.24)	0.17 (0.04)	3099 (208)	3885 (147)

the left, because the controls are right hand sided. In that figure it is clear that patient values differ significantly from the healthy hemisphere with respect to the pathological. We have also observed that there is a good correlation between the obtained values and the state of the patient, although clinical tests have to be done to validate these results.

4.2 Reproducibility Study

The reproducibility has been studied using three different data sets from the same subject, acquired with different parameters. The first volume was acquired with the same parameters as the data from section 2, in the second one we used NEX=2, and in the third data we used a matrix size of 192x192. Table 2 shows the mean and standard deviation (std) of the measures obtained in the pyramidal

tract, where the std values obtained now, almost one order of magnitude lower than in table 1, proves that our quantification method is reproducible.

5 Conclusions

We have demonstrated that our proposed methodology to measure fiber tracts is useful to differentiate between control subjects and patients. The values obtained with our new integrity, discontinuity and connectivity measures are more discriminant than usual measures such as the mean FA, or number of fibers. Also the quantification method presented here is valid to distinguish clearly between different hemispheres in the case of a neurological disorder affecting the corticospinal tract. These measures are also extensible to other main fiber bundles such as the corpus callosum, the uncinate fasciculus or the cingulum.

We have also shown the validity of the measures proposed with the reproducibility study, although further analysis have to be done using other tractography techniques, to try to deal with fiber crossings, other fiber bundles and other acquisition parameters.

References

1. Lawes, I., Barrick, T., Murugam, V., Spierings, N., Evans, D., et al.: Atlas-based segmentation of white matter tracts of the human brain using diffusion tensor tractography and comparison with classical dissection. Neuroimage 39, 62–79 (2008)
2. Hagler, D.J., Ahmadi, M., Kuperman, J., Holland, D., McDonald, C., Halgren, E., Dale, A.: Automated white matter tractography using a probabilistic diffusion tensor atlas: application to temporal lobe epilepsy. Hum. Brain Mapp. 30(5), 1535–1547 (2009)
3. Hua, K., Zhang, J., Wakana, S., Jiang, H., Li, X., Reich, D.S., Calabresi, P., Pekar, J., et al.: Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and trat-specific quantification. Neuroimage 39, 336–347 (2008)
4. Pagani, E., Filippi, M., Rocca, M., Horsfield, M.: A method for obtaining tract-specific diffusion tensor MRI measurements in the presence of disease: application to patients with clinically isolated syndromes suggestive of multiple sclerosis. NeuroImage 26, 258–265 (2005)
5. Correia, S., Lee, S., Voorn, T., Tate, D., Paul, R., Zhang, S., Salloway, S., Malloy, P., Laidlaw, D.: Quantitative tractography metrics of white matter integrity in diffusion-tensor MRI. Neuroimage 42, 568–581 (2008)
6. Smith, S., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T., Mackay, C., Watkins, K., Ciccarelli, O., et al.: Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. Neuroimage 31, 1487–1505 (2006)
7. Mori, S., Wakana, S., Nagae-Poetscher, L.M., van Zijl, P.C.M.: MRI Atlas of Human Whitte Matter. Elsevier, Amsterdam (2005)

A Riemannian Framework for Orientation Distribution Function Computing

Jian Cheng[1,2], Aurobrata Ghosh[2], Tianzi Jiang[1], and Rachid Deriche[2]

[1] LIAMA Research Center for Computational Medicine, Institute of Automation, Chinese Academy of Sciences, China
[2] Odyssée Project Team, INRIA Sophia Antipolis – Méditerranée, France
{Jian.Cheng,Aurobrata.Ghosh}@sophia.inria.fr

Abstract. Compared with Diffusion Tensor Imaging (DTI), High Angular Resolution Imaging (HARDI) can better explore the complex microstructure of white matter. Orientation Distribution Function (ODF) is used to describe the probability of the fiber direction. Fisher information metric has been constructed for probability density family in Information Geometry theory and it has been successfully applied for tensor computing in DTI. In this paper, we present a state of the art Riemannian framework for ODF computing based on Information Geometry and sparse representation of orthonormal bases. In this Riemannian framework, the exponential map, logarithmic map and geodesic have closed forms. And the weighted Frechet mean exists uniquely on this manifold. We also propose a novel scalar measurement, named Geometric Anisotropy (GA), which is the Riemannian geodesic distance between the ODF and the isotropic ODF. The Renyi entropy $H_{\frac{1}{2}}$ of the ODF can be computed from the GA. Moreover, we present an Affine-Euclidean framework and a Log-Euclidean framework so that we can work in an Euclidean space. As an application, Lagrange interpolation on ODF field is proposed based on weighted Frechet mean. We validate our methods on synthetic and real data experiments. Compared with existing Riemannian frameworks on ODF, our framework is model-free. The estimation of the parameters, i.e. Riemannian coordinates, is robust and linear. Moreover it should be noted that our theoretical results can be used for any probability density function (PDF) under an orthonormal basis representation.

1 Introduction

Diffusion Tensor Imaging (DTI) [1], based on the assumption of Gaussian diffusion, is a popular way to probe the white matter in vivo. Riemannian frameworks (Affine Invariant framework and Log-Euclidean framework) have been proposed for computing on tensor valued data [2,3,4,5]. These frameworks have proved useful for regularizing, registering, segmenting, and interpolating tensor data. The Riemannian framework provides the intrinsic way to deal with objects in a non-Euclidean space. High Angular Resolution Diffusion Imaging (HARDI) is used to probe non-Gaussian diffusion which represents more intricate microstructure in the tissue. In HARDI data modeling the role of the Orientation Distribution Function (ODF) [6] is pivotal. It is an antipodally symmetric probability density function defined on S^2 that is sufficiently smooth to be represented by an orthonormal basis, such as the High Order Tensor (HOT) basis [7],

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 911–918, 2009.

the Spherical Harmonic (SH) basis [8], or the Spherical Wavelet basis [9], using only a few coefficients.

A Riemannian framework has never been proposed to our knowledge for the orthonormal basis representation of ODFs. In [6], the author used Kullback-Leibler divergence, however, it's not a real metric. In [10], ODFs were represented in the SH basis and the authors used the $L2$ distance between the coefficient vectors as the distance between ODFs. The $L2$ distance is computationally efficient but it's an ambient distance, not a geodesic. Apart from orthonormal bases, special function mixture models have also been proposed to represent ODFs [11,12]. In the case of the von Mises-Fisher distributions, the Riemannian framework can be constructed on the mixture model [13]. However, not every ODF can be represented by these functions, as the von Mises-Fisher distributions don't form a basis. Additionally, there is no efficient method for estimating the parameters and the authors used an iterative optimization approach. Also, there is no natural way to define a metric in a multiplicative space because different spaces should have different weighted coefficients, but the authors considered the same weights in all the spaces.

In this paper, we propose a natural, computationally efficient and model-free Riemannian framework for computing on ODFs. First we study the parameter space of ODFs based on the orthonormal basis representation, and propose the geodesic, the exponential map, the logarithmic map, the weighted Frechet mean, and the Geometric Anisotropy (GA) and the Renyi entropy (RE). The GA we propose is the Riemannian geodesic to a spherical ODF, it indicates how non-spherical an ODF is. Next we also propose the Affine-Euclidean (AE) and the Log-Euclidean (LE) frameworks to be able to work in an Euclidean space. Finally we propose an interpolation method and computation of the mean ODF of a set of ODFs as example applications, based on weighted Frechet mean, and demonstrate our method on synthetic and real data.

2 Riemannian Framework for ODF

ODFs have been successfully represented in orthonormal bases. Compared with the special function mixture model in [11,12], basis representations are model-free methods because linear combination of these bases can represent all ODFs. In this section, we will construct the Riemannian framework under the orthonormal basis representation.

2.1 Parametric Family and Riemannian Framework

The ODF is a PDF denoted by $p(x), x \in \chi, \chi = S^2$. Let the square-root of the ODF be denoted by $\sqrt{p} : x \rightarrow \sqrt{p(x)}, \forall x \in \chi$. We assume a very weak and reasonable assumption that the square-root $\sqrt{p}$ can be represented by the linear combination of the first K bases, i.e. $\sqrt{p}(x) = \sum_{i=1}^{K} c_i B_i(x)$. $\{B_i(x)\}_{i \in N}$ is any orthonormal basis function set, and $\boldsymbol{c} = (c_1, c_2, ..., c_K)^T$ is called the Riemannian Coordinate. Thus $\boldsymbol{c}$ is the parameter of ODF $p(x|\boldsymbol{c})$, and we have the probability family PF_K as

$$PF_K = \left\{ p(x|\boldsymbol{c}) = \left(\sum_{i=1}^{K} c_i B_i(x) \right)^2 : \int_{\chi} p(x|\boldsymbol{c}) dx = \sum_{i=1}^{K} c_i^2 = 1, \ \sum_{i=1}^{K} c_i B_i(x) \geq 0, \ \forall x \in \chi \right\} \tag{1}$$

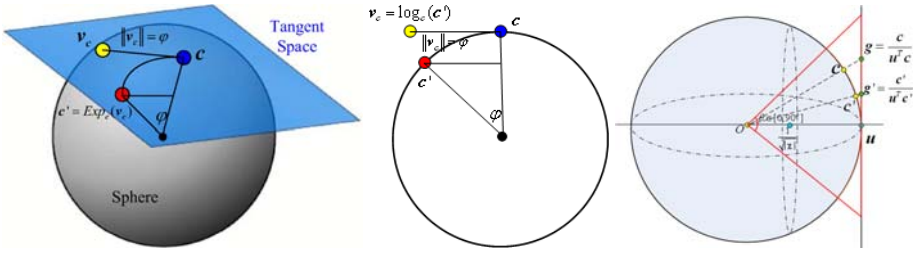


Fig. 1. *Left & Middle*: Riemannian framework. *Right*: Parameter space, Affine-Euclidean and Log-Euclidean frameworks. The isotropic ODF $\boldsymbol{u}$ is only marked as an example and may not be at the center of the cone.

All ODFs will be in PF_K if K is large enough thanks to the orthonormal basis representation. Let $\partial_i = \frac{\partial}{\partial \boldsymbol{c}_i}$, then we have the information metric as [13]:

$$g_{ij}(\boldsymbol{c}) = \int_\chi \partial_i \log\left(p(x|\boldsymbol{c})\right) \partial_j \log\left(p(x|\boldsymbol{c})\right) p(x|\boldsymbol{c}) dx = 4 \int_\chi \partial_i \sqrt{p(x|\boldsymbol{c})} \partial_j \sqrt{p(x|\boldsymbol{c})} dx \quad (2)$$

$$= 4 \int_\chi \partial_i \left(\sum_{m=1}^{K} c_m B_m(x) \right) \partial_j \left(\sum_{n=1}^{K} c_n B_n(x) \right) dx = 4 \int_\chi B_i(x) B_j(x) dx = 4\delta_{ij} \quad (3)$$

We do not care about a constant multiplier in a metric if we want to devise a computational method. So we consider $g_{ij} = \delta_{ij}$. From the formulae above, we can find that the parameter space $PS_K = \{\boldsymbol{c} \mid \|\boldsymbol{c}\| = \sum_{i=1}^{K} c_i^2 = 1,\ \sum_{i=1}^{K} c_i B_i(x) \geq 0,\ \forall x \in \chi\}$ is just a subset of the sphere S^{k-1}. The sphere is a simple manifold which has been well studied. Thus we can get the geodesic (the part of the great circle on the sphere), exponential map, and logarithmic map very easily [14,11,15]. If we let $\sqrt{p}(x|\boldsymbol{c}) = \sum_{i=1}^{K} c_i B_i(x)$, $\sqrt{p}(x|\boldsymbol{c}') = \sum_{j=1}^{K} c'_j B_j(x)$, and $\boldsymbol{v_c}$ is the tangent vector at $\boldsymbol{c}$ towards $\boldsymbol{c}'$, then we have the Riemannian framework (see Fig-1):

$$Distance:\ d_{g_{ij}}(p(\cdot|\boldsymbol{c}), p(\cdot|\boldsymbol{c}')) = d_{\delta_{ij}}(\boldsymbol{c}, \boldsymbol{c}') = \arccos\left(\sum_{i=1}^{K} c_i c'_i\right) \quad (4)$$

$$Exp:\ Exp_{\boldsymbol{c}}(\boldsymbol{v_c}) = \boldsymbol{c}' = \boldsymbol{c}\cos\varphi + \frac{\boldsymbol{v_c}}{\|\boldsymbol{v_c}\|}\sin\varphi,\ where\ \varphi = \|\boldsymbol{v_c}\| \quad (5)$$

$$Log:\ Log_{\boldsymbol{c}}(\boldsymbol{c}') = \boldsymbol{v_c} = \frac{\boldsymbol{c}' - \boldsymbol{c}\cos\varphi}{\|\boldsymbol{c}' - \boldsymbol{c}\cos\varphi\|}\varphi,\ where\ \varphi = \arccos\left(\sum_{i=1}^{K} c_i c'_i\right) \quad (6)$$

The geodesic between ODF $p(x|\boldsymbol{c})$ and ODF $p(x|\boldsymbol{c}')$ is $\gamma(t)$: $p(x|\boldsymbol{c}(t))$, where $\boldsymbol{c}(t) = Exp_{\boldsymbol{c}}(tLog_{\boldsymbol{c}}(\boldsymbol{c}'))$. Then $\gamma(0) = \boldsymbol{c}$, $\frac{d\gamma(t)}{dt}|_{t=0} = \boldsymbol{v_c}$, $\gamma(1) = \boldsymbol{c}'$. Actually, it's the part of the great circle that connects $\boldsymbol{c}$ and $\boldsymbol{c}'$.

2.2 Parameter Space, Affine-Euclidean and Log-Euclidean Frameworks

To use the Riemannian framework, we see from Equ-1 that we should keep the parameter $\boldsymbol{c}$ in PS_K. To do so we derive useful theoretical results on PS_K.

Property-1: *PS_K is closed.* The boundary of PS_K is $\{\boldsymbol{c} \mid \sum_{i=1}^{K} c_i^2 = 1,\ \exists C \subset \chi,\ s.t.\ x \in C,\ \sum_{i=1}^{K} c_i B_i(x) = 0,\ \forall x \notin C,\ \sum_{i=1}^{K} c_i B_i(x) > 0\}$. Therefore, PS_K is closed by definition.
Property-2: *PS_K is a convex subset of sphere S^{K-1}.* If $\boldsymbol{c} \in PS_K$ and $\boldsymbol{c}' \in PS_K$, let $\boldsymbol{c}(t),\ t \in [0,1]$ be the geodesic between them. Now, $\forall t \in [0,1]$, $\exists \lambda \in [0,1]$, s.t. $\boldsymbol{c}(t) = \frac{\lambda \boldsymbol{c} + (1-\lambda)\boldsymbol{c}'}{\|\lambda \boldsymbol{c} + (1-\lambda)\boldsymbol{c}'\|}$, because $\boldsymbol{c}(t)$ is on the great circle (geodesic) and $\lambda \boldsymbol{c} + (1-\lambda)\boldsymbol{c}'$ is on the chord between $\boldsymbol{c}$ and $\boldsymbol{c}'$. And $p(x|\boldsymbol{c}(t)) = \sum_{i=1}^{K} c_i(t) B(x) = \frac{\sum_{i=1}^{K} (\lambda \sqrt{p}(x|\boldsymbol{c}) + (1-\lambda) p(x|\boldsymbol{c}'))}{\|\lambda \boldsymbol{c} + (1-\lambda)\boldsymbol{c}'\|} \geq 0$.
Property-3: *PS_K is contained in a convex cone with angle 90^o.* $\forall \mathbf{c}, \mathbf{c}' \in PS_K$, $\int_{\chi} \sqrt{p}(x|\mathbf{c}) \sqrt{p}(x|\mathbf{c}')dx = \int_{\chi} \sum_{i=1}^{K} c_i B_i(x) \sum_{j=1}^{K} c_j B_j(x) dx = \sum_{i=1}^{K} c_i c_i' = \cos(\mathbf{c}, \mathbf{c}') \in [0,1]$. *That means the angle between any two points in PS_K is less than 90^o (see Fig-1).*
Property-4: *The projection of any $\boldsymbol{c} \in PS_K$ on the Riemannian Coordinate $\mathbf{u}$ of the isotropic ODF should be more than $\frac{1}{\sqrt{4\pi}}$.* If the isotropic ODF is denoted by $U(x) = \frac{1}{|\chi|} = \frac{1}{4\pi}$, and $\sqrt{U}(x) = \frac{1}{\sqrt{|\chi|}} = \sum_{i=1}^{K} u_i B_i(x)$, then $\forall \boldsymbol{c} \in PS_K$, $\boldsymbol{c}^T \boldsymbol{u} = \cos(\boldsymbol{c}, \boldsymbol{u}) = \int_{\chi} \sqrt{p}(x|\boldsymbol{c}) \frac{1}{\sqrt{|\chi|}} > \frac{1}{\sqrt{|\chi|}} \int_{\chi} p(x|\boldsymbol{c}) dx = \frac{1}{\sqrt{4\pi}}$. That means the projection of $\boldsymbol{c}$ on $\boldsymbol{u}$ should be more than $\frac{1}{\sqrt{4\pi}}$. If we choose the SH basis, $\forall \boldsymbol{c} = (c_1, c_2, ..., c_K) \in PS_K$, $c_1 > \frac{1}{\sqrt{4\pi}}$ because $\boldsymbol{u} = (1, 0, ..., 0)$.

Result-1: *Geometric Anisotropy (GA).* Thanks to property-2, in order to measure the anisotropy of an ODF, it is possible to define GA, a scalar function, as the geodesic distance between the ODF $p(x|\boldsymbol{c})$ and the isotropic ODF $U(x)$, i.e.

$$GA(p(x|\boldsymbol{c})) = d(p(\cdot|\boldsymbol{c}), U(\cdot)) = \arccos(\boldsymbol{c}^T \boldsymbol{u}) \in [0, \arccos(\frac{1}{\sqrt{4\pi}}))) \tag{7}$$

Result-2: *Renyi Entropy (RE).* The RE of order $\frac{1}{2}$ is

$$\begin{aligned} H_{\frac{1}{2}}(p(x|\boldsymbol{c})) &= 2 \log \left(\int_{\chi} \sqrt{p}(x|\boldsymbol{c}) dx \right) = 2 \log \left(\cos(GA(p(x|\boldsymbol{c}))) \right) + \log(4\pi) \\ &= \log(4\pi (\boldsymbol{c}^T \boldsymbol{u})^2) \in (0, \log(4\pi)] \end{aligned} \tag{8}$$

Result-3: *Affine-Euclidean (AE) & Log-Euclidean (LE) frameworks.* Since PS_K is a closed and convex subset of a hemisphere including $\boldsymbol{u}$ (properties-1,2,3,4), it is diffeomorphic to a subset, named *Euc*, of the tangent space of $\boldsymbol{u}$. We can define a diffeomorphism $F : PS_K \subset S^{K-1} \rightarrow Euc \subset T_{\boldsymbol{u}}$. We can use the metric in *Euc* to redefine the metric in PS_K. Then the Riemannian framework will be changed so that we can work in an Euclidean space, which will approximate the Riemannian framework (see Fig-1). If we Let F_* and F^* denote respectively the push forward and pull back maps, then

$$Distance : \ d_{PS_K}(p(\cdot|\boldsymbol{c}), p(\cdot|\boldsymbol{c}')) = d_{Euc}(F(\boldsymbol{c}), F(\boldsymbol{c}')) = \|F(\boldsymbol{c}) - F(\boldsymbol{c}')\| \tag{9}$$

$$Exp : \ Exp_{\boldsymbol{c}}(\boldsymbol{v_c}) = \boldsymbol{c}' = F^{-1}(F(\boldsymbol{c}) + F_* \boldsymbol{v_c}) \tag{10}$$

$$Log : \ Log_{\boldsymbol{c}}(\boldsymbol{c}') = \boldsymbol{v_c} = F^*(F(\boldsymbol{c}') - F(\boldsymbol{c})) \tag{11}$$

If we choose $F(\boldsymbol{c}) = Log_{\boldsymbol{u}}(\boldsymbol{c})$, then we get the Log-Euclidean framework like the one in tensor space [4]. If we choose $F(\boldsymbol{c}) = \frac{\boldsymbol{c}}{\boldsymbol{u}^T \boldsymbol{c}} - \boldsymbol{u}$, we get the Affine-Euclidean framework, which is very popular in Statistical Shape Analysis (SSA) [5,16].

2.3 Weighted Frechet Mean

Another important property of PS_K is that the weighted Frechet mean exists uniquely [14]. The weighted Frechet mean is defined as the point on the manifold PS_K that minimizes the summation of weighted squared Riemannian geodesic distances to each sample point [3]. If $f_1, f_2, ..., f_N \in PS_K$, $Q_{\boldsymbol{w}}(f) = \sum_{i=1}^{N} w_i d(f, f_i)^2$, $\boldsymbol{w} = (w_1, ..., w_N)^T$, $w_i \in [0, 1]$, $\sum_{i=1}^{N} w_i = 1$, then

$$\mu_{\boldsymbol{w}} = \arg \min_{f \in PS_K} Q_{\boldsymbol{w}}(f) = \arg \min_{f \in PS_K} \sum_{i=1}^{N} w_i d(f, f_i)^2 \tag{12}$$

$Q_{\boldsymbol{w}}(\mu_{\boldsymbol{w}})$ is called as sample weighted variance. In [14], the author proved that Frechet mean on hemisphere exists uniquely, so it does in our parameter space because $\boldsymbol{c}^T \boldsymbol{u} > \frac{1}{\sqrt{4\pi}}$. Borrowing from the notation in [14], we denote $\mu_{\boldsymbol{w}} \triangleq \widetilde{\sum}_{i=1}^{N} w_i f_i \triangleq w_1 f_1 \oplus w_2 f_2 \oplus \cdots \oplus w_N f_N$. Here is a gradient descent method to calculate the weighted Frechet mean.

Algorithm 1. Weighted Frechet Mean
Input: $f_1, ..., f_N \in PS^K$, $\boldsymbol{w} = (w_1, ..., w_N)'$, $w_i \geq 0, i = 1, 2, ..., N$, $\sum_{i=1}^{N} w_i = 1$.
Output: $\mu_{\boldsymbol{w}}$, the *Weighted Frechet Mean*.

Initialization: $\mu_{\boldsymbol{w}}^{(0)} = \frac{\sum_{i=1}^{N} w_i f_i}{\| \sum_{i=1}^{N} w_i f_i \|}$, $k = 0$
Do
$v_{\mu_{\boldsymbol{w}}^{(k)}} = \sum_{i=1}^{N} w_i Log_{\mu_{\boldsymbol{w}}^{(k)}}(f_i)$
$\mu_{\boldsymbol{w}}^{(k+1)} = Exp_{\mu_{\boldsymbol{w}}^{(k)}}(v_{\mu_{\boldsymbol{w}}^{(k)}})$
$k = k + 1$
while $\|v_{\mu_{\boldsymbol{w}}^{(k)}}\| > \varepsilon$

For AE and LE frameworks, because we use the distance on Euclidean spaces to approximate the distance on PS, obviously there is a closed form:

$$\mu_{\boldsymbol{w}} = F^{-1}\left(\sum_{i=1}^{N} w_i F(f_i) \right) \tag{13}$$

3 Riemannian Coordinate Estimation and ODF Interpolation

In the previous section, our theoretical results are independent of the choice of the basis function, and we assume that for every ODF $p(x|\boldsymbol{c})$, we know the Riemannian Coordinate $\boldsymbol{c}$. Here we show a simple and linear method to estimate it from raw HARDI data under the SH basis representation.

First, we can use any method to estimate the ODF $p(x)$. Here we choose the robust, linear, analytical Q-ball imaging [8] with the regularization term. *Second*, we take discrete samples $p(x_i)$ of the ODF on the continuous sphere, and calculate the square-root of these samples $\sqrt{p(x_i)}$. *Third*, we use these square-root samples to estimate the square-root of the ODF through a least squares SH basis fit linearly without regularization. If we choose the first N order SH bases in the first step ($N \geq 4$) [8], and the first M order in the third step, we should keep $M \geq N$. The advantages of our estimation method are its linearity, robustness and speed, thanks to the analytical Q-ball imaging.

Based on the weighted Frechet mean in section 2.3, we propose a Lagrange interpolation method on ODF space as an application. In 1-dimension case, if we have N ODFs $\{f(x_i),\ i = 1, 2, ..., N\}$ in N spatial position $\{x_i, i = 1, 2, ..., N\}$. , we can interpolate the ODF in the whole spatial axis as the weighted Frechet mean of $\{f(x_i)\}$, i.e. $f(x) = \widetilde{\sum}_{i=1}^{N} w_i(x) f(x_i)$, $w_i(x) = \prod_{l=1, l\neq i}^{N} \frac{x-x_l}{x_i-x_l}$. It should be noted that if $N = 2$, then $f(x) = (1-t)f(x_1) \oplus tf(x_2)$, $t = \frac{x-x_1}{x_2-x_1}$. Actually, in that case, $f(x)$ is the geodesic determined by $f(x_1)$ and $f(x_2)$, i.e. $f(x) = Exp_{f(x_1)}\left(tLog_{f(x_1)}(f(x_2))\right)$. For HARDI image in 3-dimension, if we have N_1, N_2 and N_3 for ODFs for each dimension, i.e. $\{f(x_i, y_j, z_k),\ 1 \le i \le N_1, 1 \le j \le N_2, 1 \le k \le N_3\}$, we can interpolate ODF in the continuous 3D space as the weighted Frechet mean of $\{f(x_i, y_j, z_k)\}$.

$$f(x, y, z) = \widetilde{\sum_{i=1}^{N_1}} \widetilde{\sum_{j=1}^{N_2}} \widetilde{\sum_{k=1}^{N_3}} w_i^{(1)}(x) w_j^{(2)}(y) w_k^{(3)}(z) f(x_i, y_j, z_k) \tag{14}$$

$$w_i^{(1)}(x) = \prod_{l=1, l\neq i}^{N_1} \frac{x - x_l}{x_i - x_l},\ w_j^{(2)}(y) = \prod_{m=1, m\neq j}^{N_2} \frac{y - y_m}{y_j - y_m},\ w_k^{(3)}(z) = \prod_{n=1, n\neq k}^{N_3} \frac{z - z_n}{z_k - z_n} \tag{15}$$

4 Experiments and Results

Experiments were conducted to illustrate the proposed metrics. Synthetic ODFs were generated using the multi-tensor model (1 or 2 fibers, every tensor with eigenvalues $[1700, 300, 300] \times 10^{-6} mm^2/s$) [17]. Riemannian Coordinates of ODFs were estimated using the method proposed in section-3. In the experiments, we chose order 4 to estimate the ODF and its square-root with satisfactory results. Then the Lagrange interpolation method was applied to interpolate the ODFs in 1D or 2D cases.

GA was calculated for synthetic data and clinical real data, and compared with GFA [6]. No normalization was used for ODF visualization. Since RE $H_{\frac{1}{2}} \in (0, \log(4\pi)]$, we plotted its normalized version $H_{\frac{1}{2}} / \log(4\pi)$. From the results, we find that the three proposed metrics – AE, LE, Riemannian (RM), give very similar results. The RE is close to 1. This is because the ODF proposed in [6] is very smooth and close to the isotropic ODF. The results of the experiments-A to E are given in the Fig-2.

Exp-A: We interpolate from the isotropic ODF $U(x)$ to a single fiber using the (1)AE, (2) LE, (3) RM metrics. As expected, the GA is linear, since the GA is the Riemannian distance to $\boldsymbol{u}$. *Exp-B:* We illustrate Riemannian geodesic interpolation between different fiber configurations, and plot the corresponding GA. *Exp-C:* 2D Lagrange interpolation using the (1)AE, (2) LE, (3) RM metrics. The background is the GA. *Exps-A,B,C:* ODFs coloured by GA. *Exp-D:* GA (left) from clinical real data compared to GFA (right). ODFs estimated using [8]. *Exp-E:* Mean ODF computation from a field of ODFs using the (1) LE and (2) RM metrics.

5 Conclusion

In this paper, we proposed a novel Riemannian framework based on Information Geometry and orthonormal basis representation. Our framework is model-free and computationally efficient. The estimation method of Riemannian Coordinates is fast, linear and

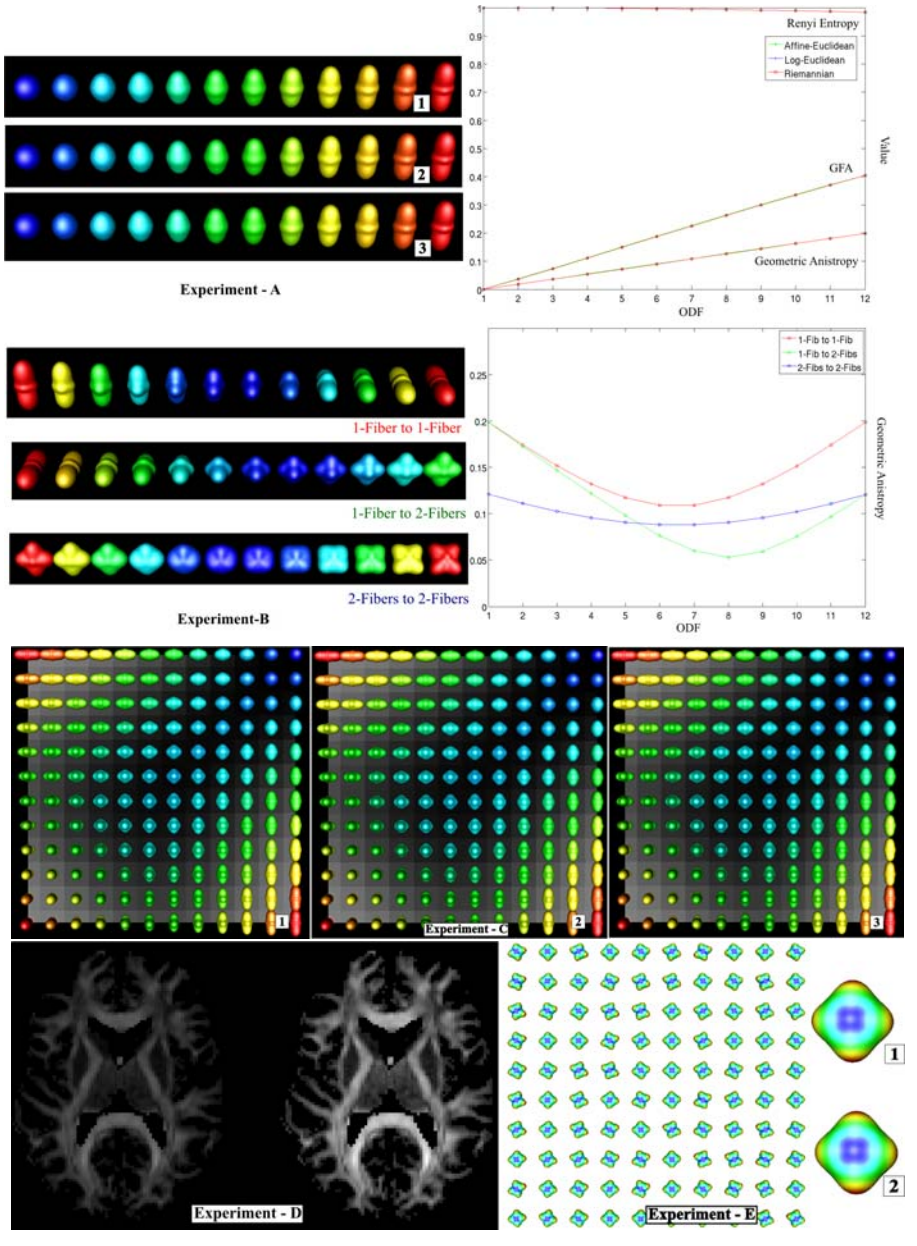


Fig. 2. *Exp-A:* Interpolation from Isotropic to 1-fiber using (1) AE, (2) LE, (3) RM metrics. Riemannian (RM) GA is linear as expected. *Exp-B:* Various RM geodesics with corresponding GA plotted. *Exp-C:* 2D Lagrange interpolation (given 4 ODFs in the 4 corners) using (1) AE, (2) LE, (3) RM metrics. Background is GA. *Exp-D:* GA (left) and GFA (right) from clinical real data. ODFs estimated using [8] *Exp-E:* (1) LE & (2) RM mean from a field of ODFs. Results and computation time of the metrics are comparable. For mean computation, LE has a closed form.

robust. The weighted Frechet mean exists uniquely. Moreover, we present the AE and the LE frameworks so that the computation is simplified further and we have closed forms. A Lagrange interpolation method was proposed as an application. The results validate our framework. Last but not least, this framework can be used for any kind of PDF under orthonormal basis representation, which is part of our future plans.

References

1. Basser, P.J., Mattiello, J., LeBihan, D.: Mr. diffusion tensor spectroscropy and imaging. Biophysical Journal 66, 259–267 (1994)
2. Pennec, X., Fillard, P., Ayache, N.: A riemannian framework for tensor computing. International Journal of Computer Vision 66, 41–66 (2006)
3. Lenglet, C., Rousson, M., Deriche, R.: Statistics on the manifold of multivariate normal distributions theory and application to diffusion tensor mri processingy. Journal of Mathematical Imaging and Vision 25(3), 423 (2006)
4. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-euclidean metrics for fast and simple calculus on diffusion tensors. Magnetic Resonance in Medicine 56, 411–421 (2006)
5. Fletcher, P.T.: Statistical Variability in Nonlinear Spaces Application to Shape Analysis and DT-MRI. PhD thesis, University of North Carolina (2004)
6. Tuch, D.S.: Q-ball imaging. Magnetic Resonance in Medicine 52, 1358–1372 (2004)
7. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Apparent diffusion coefficients from high angular resolution diffusion images: Estimation and application. Magnetic Resonance in Medicine 56, 395–410 (2006)
8. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast and robust analytical q-ball imaging. Magnetic Resonance in Medicine 58, 497–510 (2007)
9. Kezele, I., Descoteaux, M., Poupon, C., Abrial, P., Poupon, F., Mangin, J.F.: Multiresolution decomposition of hardi and odf profiles using spherical wavelets. In: MICCAI (2008)
10. Descoteaux, M., Deriche, R.: High angular resolution diffusion mri segmentation using region-based statistical surface evolution. Journal of Mathematical Imaging in Vision special issue on Mathematics in Image Analysis, 305–317 (2008)
11. McGraw, T., Vemuri, B.C., Yezierski, R., Mareci, T.H.: Segmentation of high angular resolution diffusion MRI modeled as a field of von mises-fisher mixtures. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 463–475. Springer, Heidelberg (2006)
12. Rathi, Y., Michailovich, O., Shenton, M.E., Bouix, S.: Directional functions for orientation distribution estimation. Medical Image Analysis, 432–444 (2009)
13. Amari, S., Nagaoka, H.: Methods of Information Geometry. American Mathematical Society, USA (2000)
14. Buss, S.R., Fillmore, J.P.: Spherical averages and applications to spherical splines and interpolation. ACM Transactions on Graphics 20, 95–126 (2001)
15. Srivastava, A., Jermyn, I.: Riemannian analysis of probability density functions with applications in vision. In: CVPR (2007)
16. Dryden, I.L., Mardia, K.V.: Statistical Shape Analysis. John Wiley, Chichester (1998)
17. Descoteaux, M.: High Angular Resolution Diffusion MRI: from Local Estimation to Segmentation and Tractograpghy. PhD thesis, INRIA Sophia Antipolis (2008)

Bias of Least Squares Approaches for Diffusion Tensor Estimation from Array Coils in DT–MRI

Antonio Tristán-Vega[1], Carl-Fredrik Westin[2], and Santiago Aja-Fernández[1]

[1] Laboratory of Image Processing. Universidad de Valladolid, Spain
[2] LMI. Brigham and Women's Hospital, Boston
atriveg@lpi.tel.uva.es, westin@bwh.harvard.edu, sanaja@tel.uva.es

Abstract. Least Squares (LS) and its weighted version are standard techniques to estimate the Diffusion Tensor (DT) from Diffusion Weighted Images (DWI). They require to linearize the problem by computing the logarithm of the DWI. For the single-coil Rician noise model it has been shown that this model does not introduce a significant bias, but for multiple array coils and parallel imaging, the noise cannot longer be modeled as Rician. As a result the validity of LS approaches is not assured. An analytical study of noise statistics for a multiple coil system is carried out, together with the Weighted LS formulation and noise analysis for this model. Results show that the bias in the computation of the components of the DT may be comparable to their variance in many cases, stressing the importance of unbiased filtering previous to DT estimation.

1 Introduction

Least Squares (LS) techniques have become the standard to estimate the Diffusion Tensor (DT) in DT–MRI [1]. Although other approaches are possible [2], LS are fast and robust, and they show optimal properties (minimum variance) when Weighted Least Squares (WLS) is used [1]. For single-coil acquisition systems where the noise is Rician distributed [3], the estimation is nearly unbiased, so WLS is in this case the Best Linear Unbiased Estimator (BLUE). With the advent of multiple receiver coils for similtaneous acquisition and parallel reconstruction schemes (pMRI), the noise in the Diffusion Weighted Images (DWI) is no longer Rician [4,5,6] and the properties of WLS have not been analyzed for these composite signals. In addition, although the importance of the bias in DWI has been studied in filtering problems, [7,8,9,10], the necessity of bias removal for DT estimation has not been either thoroughly analyzed to the very best of our knowledge. Therefore the main contributions of the present paper are: **first**, we theoretically justify the conclusions of [1] regarding the bias and variance of log–Rician distributed signals; for array coils, the log–non–central χ model has been suggested [5], so we derive both analytical expressions and practical approximations for its statistical characterization. Based on this analysis, we **secondly** extend the work in [1] to multiple coils DT–MRI, showing that it admits the same WLS formulation. **Third** we show that, as opposed to the case of Rician noise, the bias in WLS estimation is relevant with array coils as

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 919–926, 2009.

a consequence of the estimation of the composite magnitude signal by means of Sum of Squares (SoS). This implies that the analysis may be generalized to some pMRI reconstruction schemes based on SoS, such as the popular GRAPPA technique. At the same time, the need for unbiased filtering before DT estimation is evidenced.

2 Statistics of Noise in the Log–Domain

Under the assumption of narrow pulses, and for mono–compartment Gaussian diffusion, the tensor model relates the DWI signal A_i for each gradient direction i to the T2 baseline image A_0 following the Stejskal–Tanner equation [11]:

$$A_i = A_0 \exp(-b g_i^T D g_i) \Leftrightarrow \log(A_0) - \log(A_i) = b g_i^T D g_i,\ 1 \leq i \leq N \quad (1)$$

where g_i are the N gradient directions, b is the diffusion weighting parameter, and D is the DT. Assuming A_i real, for a single–coil acquisition the received signal M_i is the Rician distributed envelope of a complex Gaussian process [3]:

$$M_i = \sqrt{\left(A_i + n_{c,i}^2\right)^2 + n_{s,i}^2} \quad (2)$$

where $n_{c,i}$ and $n_{s,i}$ are independent Gaussian processes with zero mean and standard deviation σ. In phased array systems with multiple coils –assuming no subsampling in the **k**-space–, the composite magnitude signal after sum-of-squares combination may be modeled as [4,5]:

$$M_i = \sqrt{\sum_{l=1}^{L} \left(A_{i,l} + n_{c,i,l}^2\right)^2 + n_{s,i,l}^2} \quad (3)$$

for L receiving coils. This correponds to a non–central χ distribution. Note that the Rician distribution is a particular case for $L = 1$, so the latter may be seen as a general model for single and multiple coils acquisition schemes. However, as shown in Section 3, the statistics of interest are those related to the logarithm of M; the following expressions for the mean and variance may be proved[1]:

$$E\{\log(M)\} = \frac{1}{2}\log\left(2\sigma_n^2\right) + \frac{1}{2}\frac{a}{L}{}_2F_2\left(1, 1:2, 1+L; -a\right) + \frac{1}{2}\psi(L); \quad (4)$$

$$\mathrm{Var}\{\log(M)\} = \frac{1}{4}\left[\widetilde{\mathrm{N}}_L(a) - 2\log(2\sigma^2)\frac{a}{L}{}_2F_2(1, 1:2, 1+L; -a) - \left(\psi(L) + \frac{a}{L}{}_2F_2\left(1, 1:2, 1+L; -a\right)\right)^2\right]; \quad (5)$$

with:

$$\widetilde{\mathrm{N}}_k(x) = e^{-x}\sum_{p=0}^{\infty}\frac{x^p}{\Gamma(p+1)}\left[\left(\psi(p+k)\right)^2 + \psi^{1)}(p+k)\right], \quad (6)$$

[1] We have to omit the detailed proof here due to space constraints.

where $\psi(x)$ is the polygamma function and $\psi^{1)}(x)$ is its first derivative. For convenience we have dropped the i subindex and used the notation $A_L^2 = \sum_{l=1}^{L} A_l^2$ and $a = A_L^2/2\sigma^2$. Eq. (4) generalizes the expression given in [1, eq. (11)] for the mean of the log–Rician distribution, while eq. (5) has not been previously reported to the best of our knowledge. The complexity of these expressions does not allow a detailed analysis, so we use practical approximations. To do so, let us compute the Taylor series expansion of $\log(M)$ for $\sigma \ll A_L$; without loss of generality, we may assume that $A_L^2 = LA^2$, and then:

$$\log(M) = \log(A_L) + \frac{1}{AL}\sum_{i=1}^{L} n_{c,l} + \frac{1}{2A^2L^2}\sum_{k=1}^{L}\sum_{l=1}^{L}(L\delta_{kl} - 2)\, n_{c,k} n_{c,l} + \frac{1}{2LA^2}\sum_{l=1}^{L} n_{s,l}^2 + \mathcal{O}\left(\frac{\sigma^3}{A^3}\right) \tag{7}$$

with δ the Kronecker delta function. From eq. (7) it follows:

$$E\{\log(M)\} = \log(A_L) + (L-1)\frac{\sigma^2}{A_L^2} + \mathcal{O}\left(\frac{\sigma^3}{A^3}\right) \tag{8}$$

In fact, since the expectation of all terms in $\mathcal{O}(\sigma^3/A^3)$ represent odd–order moments of Gaussians, we could have written $\mathcal{O}(\sigma^4/A^4)$ instead. The Mean Squared Error (MSE) in the estimation may be computed as the variance plus the squared bias, so we must compute the expansion of the variance to order 4 (comparable to $\text{bias}^2 = \mathcal{O}(\sigma^4/A_L^4) \equiv \mathcal{O}(a^{-2})$), being necessary to compute the expansion of $\log(M)$ to order 3. After some calculations, it yields:

$$\text{Var}\{\log(M)\} = \frac{1}{2}a^{-1} - \frac{3L-4}{4}a^{-2} + \mathcal{O}\left(a^{-3}\right) \tag{9}$$

$$\text{bias}^2\{\log(M)\} = \frac{(L-1)^2}{4}a^{-2} + \mathcal{O}\left(a^{-3}\right) \tag{10}$$

For the Rician case ($L = 1$), the squared bias is in the order of $\mathcal{O}(a^{-3}) = \mathcal{O}(\sigma^6/A_L^6)$, so it is not so important, as has been noted in [1]. For $L > 1$, the bias may be important if L is in the same order of magnitude as A_L/σ.

3 Tensor Fitting Based on Weighted Least Squares

From the linearized version of the Stejskal–Tanner model in eq. (1), the estimation of the 6 free coefficients of the DT may be seen as a WLS problem [1]:

$$\log(A_0) - \log(M_i) = \left[g_{i,1}^2,\ 2g_{i,1}g_{i,2},\ 2g_{i,1}g_{i,3},\ g_{i,2}^2,\ 2g_{i,2}g_{i,3},\ g_{i,3}^2\right] \cdot \left[bD_{11},\ bD_{12},\ bD_{13}, bD_{22},\ bD_{23},\ bD_{33}\right]^T + \epsilon_i \tag{11}$$

As opposed to [1], we do not include A_0 in the estimation, i.e., we consider that A_0 is known without any uncertainty (A_0=M_0); since the baseline is generally far

less noisy than the gradient images (its value of a is much higher), and moreover several baseline images are often available, the error due to A_0 is negligible and therefore it does not affect our analysis. Nevertheless, a similar analysis to that here presented may be carried out including A_0 in the estimation with almost identical conclusions. The WLS problem may then be written:

$$G^T W \left(\mathbf{Y} - G\mathbf{X}\right) = 0 \Rightarrow \mathbf{X} = \left(G^T W G\right)^{-1} G^T W \mathbf{Y} \tag{12}$$

where $\mathbf{Y}$ is an $N \times 1$ vector representing each of the DWI data $\log(M_i) - \log(A_0)$, $\mathbf{X}$ is the 6×1 vector of unknowns (bD_{ij}), G is the $N \times 6$ matrix resulting from the concatenation of each row in eq. (11), $[g_{i,1}^2, \ldots, g_{i,3}^2]$, and W is the $N \times N$ matrix of weights. The Gauss–Markov theorem states that under very weak assumptions the WLS is the minimum variance estimate if W is chosen to be the inverse of the covariance matrix of data, $C_{\mathbf{YY}}$. If the estimator is unbiased (as is approximately the case for Rician noise), WLS is in fact the BLUE. Since the noise in each gradient image is assumed to be independent, $C_{\mathbf{YY}}$ reduces to a diagonal matrix with diagonal elements equal to the variance of each term $\log(M_i) - \log(A_0)$, which is the same as the variance of $\log(M_i)$ since we consider A_0 known. Therefore, $W = \mathrm{diag}(W_{ii})$ with:

$$W_{ii} = \mathrm{Var}^{-1}\left\{\mathbf{Y}_i\right\} \simeq \frac{1}{a_i^{-1}/2 - a_i^{-2}\left(3L-4\right)/4} \simeq 2a_i + (3L-4) \tag{13}$$

However, we are going to neglect the terms of order a_i^{-2}, so that we will fix $W_{ii} = 2a_i = A_{Li}^2/\sigma^2$. We do so for two reasons:

1. This formulation is identical to the traditional WLS for Rician noise, see [1], and is the one implemented in all the software tools for DT estimation. Note that even in the case of Rician noise the term in a^{-2} is not null (but very small), and even so the DT is always estimated with this formulation.
2. The weights W_{ii} in fact have to be only proportional to $2a_i$, so with this formulation it is not necessary to know the exact value of σ, since we may take any weight proportional to A_{Li}^2; if we include the term in a^{-2}, it is necessary to estimate σ at each image location and to include the number of coils, L, as an additional parameter of the algorithm.

Once again, the whole analysis may be performed with the exact expression for W_{ii} with very similar conclusions. Finally, note that the weights W_{ii} depend on A_{Li} (not M_i), so they cannot be known *a priori*; this pitfall may be obviated by iteratively estimating $\mathbf{X}$ and updating the value of W [1]. Therefore, in what follows we will consider that W_{ii} are known without uncertainty.

4 Variance and Bias of the Tensor Components

In what follows, we are going to call $\mathcal{L} = (G^T W G)^{-1}$. Then, from eqs. (9), (10) and (12), and with our election of W_{ii} the covariance matrix of $\mathbf{X}$ is:

$$C_{\mathbf{XX}} = E\left\{\left(\mathbf{X} - E\{\mathbf{X}\}\right)\left(\mathbf{X} - E\{\mathbf{X}\}\right)^T\right\} = \mathcal{L}G^T W C_{\mathbf{YY}} W^T G \mathcal{L}^T$$

$$= \mathcal{L}G^T W \operatorname{diag}\left(\frac{1}{2}a_i^{-1} - \frac{\Theta(L)}{4}a_i^{-2}\right) WG\mathcal{L}$$
$$= \mathcal{L}G^T WG\mathcal{L} - \frac{\Theta(L)}{4}\mathcal{L}G^T W \operatorname{diag}\left(a_i^{-2}\right) WG\mathcal{L} = \mathcal{L} - \Theta(L)\mathcal{L}G^T G\mathcal{L} \quad (14)$$

where $\Theta(L) = 3L-4$ and W, $\mathcal{L}$ are symmetric. Note that the second term in eq. (14) appears due to the fact that we are using $W_{ii} = 2a_i$ instead of the *true* value of $\mathrm{Var}^{-1}\{Y_i\}$, and it is not null even for the Rician case (yet, $\Theta(L)$ is negative for $L = 1$, so for the Rician case the variance is greater than $\mathcal{L}$, contrary to the case of $L > 1$). For the bias, since $\mathrm{bias}\{\mathbf{Y}_i\} = a_i^{-1}(L-1)/2$, we have:

$$\mathrm{bias}\{\mathbf{X}\} = (G^T WG)^{-1} G^T W \mathrm{bias}\{\mathbf{Y}\} = (L-1)\mathcal{L}G^T \mathbf{e}_1 \quad (15)$$

where $\mathbf{e}_1$ is an $N \times 1$ vector of ones. For the 6×1 vector $G^T\mathbf{e}_1$, let us assume that the gradient directions are uniformly distributed on the sphere; each component g_k, $k = 1,2,3$ may be seen as a uniform random variable, $g_k \sim U(-1,1)$, so that $\sum_{i=1}^N g_{i,k}^2 \equiv NE\{g_k^2\} = N/3$, $\sum_{i=1}^N 2g_{i,j}g_{i,k} \equiv 2NE\{g_j\}E\{g_k\} = 0$ and $G^T\mathbf{e}_1 = [N/3, 0, 0, N/3, 0, N/3]^T = N\nu$. Although this reasoning may seem quite naive, we have checked this result to be extremely accurate for all N in the case of antipodal symmetric gradients. Finally, we define the MSE as:

$$\mathrm{MSE} = \mathrm{Var}\{\mathbf{X}\} + \mathrm{bias}^2\{\mathbf{X}\} = b^2\left(\Delta_{11}^2 + \Delta_{12}^2 + \Delta_{13}^2 + \Delta_{22}^2 + \Delta_{23}^2 + \Delta_{33}^2\right)$$
$$= \mathrm{tr}\left(\mathcal{L}\right) - \Theta(L)\mathrm{tr}\left(\mathcal{L}G^T G\mathcal{L}\right) + (L-1)^2 N^2 \nu^T \mathcal{L}^2 \nu \quad (16)$$

for $\Delta_{ij} = D_{ij} - \widetilde{D}_{ij}$ and $\widetilde{D}_{ij}$ the WLS estimate of D_{ij}.

5 Results and Discussion

To begin with, we will verify the accuracy of the approximations given by eqs. (9) and (10): in Fig. 1 (left) we represent the true values given by eqs. (4) and (5), together with empirically obtained values (for 2000 independent samples for each a and L), superimposed to the proposed approximations. As may be seen, these approximations are very accurate except for very low (unrealistic) values of a; the larger L, the larger the value of a below which the approximation diverges from the actual value. Nevertheless, note that the difference is noticeable only for $L = 8$ (below $a = 60$). As shown later on (and in Fig. 1, right) for $L = 8$ the values of interest for a are in the order of 140; for larger values of L, the approximations are inexact for higher a, but at the same time the study of the bias is of interest for higher a, so our approximation is good enough in all cases.

We now assess the impact of the bias compared to the variance in the components of the DT. First, let us assume a simplified scenario where all W_{ii} are equal to $2a_i = 2a$; then the MSE reduces to:

$$\mathrm{MSE} = \left(\frac{a^{-1}}{2} - \frac{a^{-2}}{4}\Theta(L)\right)\mathrm{tr}\left(\left(G^T G\right)^{-1}\right) + a^{-2}\frac{(L-1)^2 N^2}{4}\nu^T\left(G^T G\right)^{-2}\nu \quad (17)$$

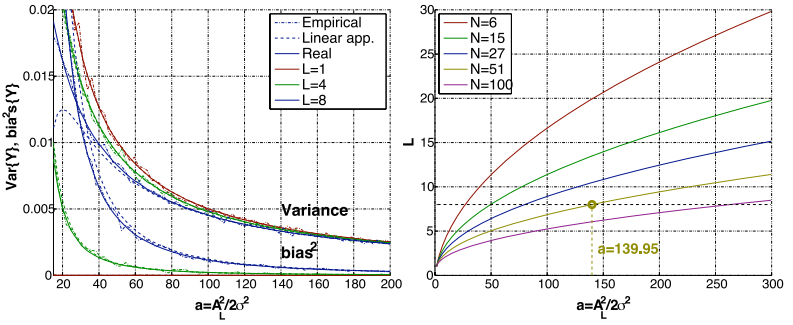


Fig. 1. Left: bias and variance in the DWI signals as a function of a for different numbers of coils; we represent true and empirically computed values together with our approximations. **Right**: minimum number of receiving coils required (for each a and N) so that the (squared) bias equals the variance in the DT components (see eq. (18)).

since $\mathcal{L} = a^{-1}/2(G^TG)^{-1}$. Therefore it is interesting to study the behavior of G^TG; a similar reasoning to that made for the calculation of ν shows that it always has the eigenvector $\nu/\|\nu\|$ associated to the eigenvalue $N/3$ for antipodal symmetric gradient directions. On the other hand, we have empirically tested that the approximation: $\mathrm{tr}((G^TG)^{-1}) \simeq 29.3/N$ is very accurate for all values of N. Therefore, we may write:

$$\begin{aligned}\mathrm{MSE} &\simeq \frac{29.3}{N}\left(\frac{a^{-1}}{2} - \frac{a^{-2}}{4}(3L-4)\right) + a^{-2}\frac{9(L-1)^2N^2}{4N^2}\nu^T\frac{\nu\nu^T}{\|\nu\|^2}\nu \\ &= \frac{29.3}{N}\left(\frac{a^{-1}}{2} - \frac{a^{-2}}{4}(3L-4)\right) + a^{-2}\frac{3(L-1)^2}{4} \end{aligned} \tag{18}$$

since $(G^TG)^{-2}$ has the eigenvector $\nu/\|\nu\|$ associated to $9/N^2$ and all other eigenvectors are orthogonal to ν. While the variance diminishes with the number of gradients, this is not the case for the bias: in Fig. 1 (right) we show the number of coils needed for each a and N so that the term corresponding to the bias in eq. (18) equals the term corresponding to the variance. For example, with $L = 8$ coils and 51 gradient directions, it is shown that if $a < 139.95$ the bias is more important than the variance. Alternatively, for $a < 139.95$ and $L = 8$ it makes little sense to take more than 51 gradients, since the variance is reduced but not the bias, so the acquisition time is increased unfruitfully; for $a = 50$, $L = 8$, the situation is the same for more than 15 gradients. Therefore, it may be more convenient to increase the number of repetitions (NEX) in the acquisition; this has the effect of dividing the noise power σ^2 by NEX (since the signals are averaged in the $\mathbf{k}$–space, where the noise is Gaussian distributed), so that:

$$\mathrm{MSE} = \frac{29.3}{N\cdot\mathrm{NEX}}\left(\frac{a^{-1}}{2} - \frac{a^{-2}}{4\ \mathrm{NEX}}(3L-4)\right) + a^{-2}\frac{3(L-1)^2}{4\ \mathrm{NEX}^2} \tag{19}$$

A constant value of $N \cdot \mathrm{NEX}$ takes a constant amount of acquisition time, and besides provides a constant reduction factor for the variance, but the bias

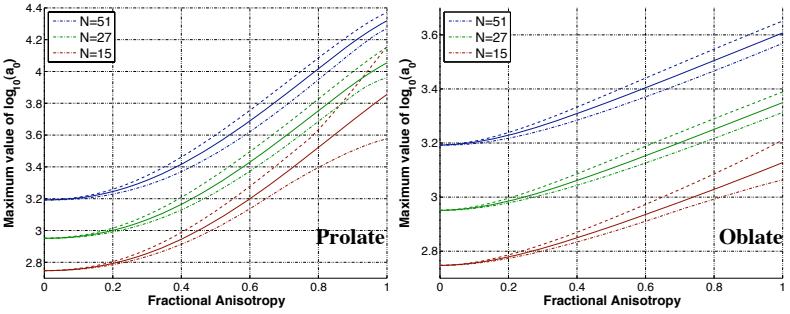


Fig. 2. Log–plot of the maximum value of $a_0 = LA_0^2/2\sigma^2$ which makes the squared bias equal to the variance in the components of the DT for different tensor shapes. We show minimum, mean and maximum values among all possible tensor orientations.

depends only on NEX. We may conclude that, for a given combination of a, N and L which makes the squared bias similar to the variance (see Fig. 1), it is preferable to increase NEX than using more gradient directions.

As a final experiment, we test the behavior when the weigths W_{ii} are not equal (non–isotropic diffusion). To do so, we use typical values $L = 8$, $N = 15,\ 27,$ and 51, $b = 1500\text{s/mm}^2$, and Mean Diffusivity (MD) $0.8 \cdot 10^{-3}\text{mm}^2/\text{s}$, and compute the maximum value of a_0 (defined as $a_0 = A_{L0}^2/2\sigma^2$ with A_{L0} the amplitude of the baseline image) for which the contribution of the bias is greater or equal than the contribution of the variance, using eq. (16). Since this result depends on the orientation of the DT, in Fig. 2 we show maximum, minimum and mean values of a_0 for different Fractional Anisotropies (FA). Two cases are considered: a prolate tensor ($\lambda_1 > \lambda_2 = \lambda_3$) and an oblate tensor ($\lambda_1 = \lambda_2 > \lambda_3$). For FA=0 (isotropic diffusion, equal weights), we have the same case as in Fig. 1 (right). With non–isotropic tensors the bias becomes even more important (a_0 may be one order of magnitude over the value for FA=0 with the prolate tensor), so the conclusions arised from our previous experiment are even reinforced. For example, with 51 directions and a prolate tensor with FA=0.8, even for $a_0 = 10^4$ ($A_{L0}/\sigma \simeq 140$) the bias is as important as the variance; although this corresponds to the baseline image, which typically shows a high SNR, it is evident that this value may be found in many image voxels. Besides, for $A_{L0}/\sigma \simeq 300$ (this SNR is an upper limit for a realistic DWI data set) the bias will be roughly 1/4 of the variance, which is clearly not negligible.

6 Conclusions

We have shown that the impact of the bias in Rician signals for WLS tensor–fitting may be neglected in any realistic case; on the contrary, for non–central χ distributed signals, the bias may be an important source of error, the larger the number of receiving coils L the more critical. While the variance in the estimation may be reduced increasing the number of gradient directions, this is not

the case for the bias. In some cases, increasing the number of gradients will not improve the estimation, since the main source of error will be the bias and not the variance. In these cases, it may be preferable to improve the SNR by increasing NEX (unless a minimum number of gradient directions is needed for some purpose). As an additional difficulty, we have shown that the traditional WLS approach is not optimal for non–central χ signals, since the weights commonly used are not those yielding minimum variance; although we have proposed a modification to avoid this problem, it makes necessary to characterize the noise power for all image voxels. Summarizing, the importance of the bias in non–central χ distributed DWI has been stated; this kind of statistics are usually found in multiple receiving coils systems with SoS reconstruction, but may be used as well to model some pMRI algortihms like GRAPPA. Therefore, an adequate unbiased filtering scheme (or a high NEX, increasing the acquisition time) should be implemented before DT estimation with modern DT–MRI protocols.

Acknowledgments. Work partially funded by grant numbers TEC2007–67073/TCM from the CICyT (Spain) and NIH R01 MH074794, NIH P41 RR13218.

References

1. Salvador, R., Pea, A., Menon, D.K., Carpenter, T.A., Pickard, J.D., Bullmore, E.T.: Formal characterization and extension of the linearized diffusion tensor model. Human Brain Mapping 24, 144–155 (2005)
2. Fillard, P., Pennec, X., Arsigny, V., Ayache, N.: Clinical DT–MRI estimation, smoothing, and fiber tracking with log–euclidean metrics. IEEE Trans. Med. Imaging 26(11), 1472–1482 (2007)
3. Gudbjartsson, H., Patz, S.: The Rician distribution of noisy MRI data. Magn. Reson. Med. 34, 910–914 (1995)
4. Roemer, P., Edelstein, W., Hayes, C., Souza, S., Mueller, O.: The NMR phased array. Magn. Reson. Med. 16, 192–225 (1990)
5. Constantinides, C., Atalar, E., McVeigh, E.: Signal-to-noise measurements in magnitude images from NMR phased arrays. Magn. Reson. Med. 38, 852–857 (1997)
6. Gilbert, G., Simard, D., Beaudoin, G.: Impact of an improved combination of signal from array coils in diffusion tensor imaging. IEEE Trans. Med. Imaging 26(11), 1428–1436 (2007)
7. Aja-Fernández, S., Alberola-López, C., Westin, C.-F.: Signal LMMSE estimation from multiple samples in MRI and DT-MRI. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 368–375. Springer, Heidelberg (2007)
8. Basu, S., Fletcher, T., Whitaker, R.T.: Rician noise removal in diffusion tensor MRI. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 117–125. Springer, Heidelberg (2006)
9. Descoteaux, M., Wiest-Daesslé, N., Prima, S., Barillot, C., Deriche, R.: Impact of rician adapted non-local means filtering on HARDI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 122–130. Springer, Heidelberg (2008)
10. Wiest-Daesslé, N., Prima, S., Coupé, P., Morrissey, S.P., Barillot, C.: Rician noise removal by non-local means filtering for low signal-to-noise ratio MRI: Applications to DT-MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 171–179. Springer, Heidelberg (2008)
11. Stejskal, E.O., Tanner, J.E.: Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. J. of Chem. Phys. 42, 288–292 (1965)

A Novel Global Tractography Algorithm Based on an Adaptive Spin Glass Model

Pierre Fillard, Cyril Poupon, and Jean-François Mangin

LNAO - CEA/DSV/I^{2}BM/Neurospin, Paris, France

Abstract. This paper introduces a novel framework for global diffusion MRI tractography inspired from a spin glass model. The entire white matter fascicle map is parameterized by pieces of fibers called spins. Spins are encouraged to move and rotate to align with the main fiber directions, and to assemble into longer chains of low curvature. Moreover, they have the ability to adapt their quantity in regions where the spin concentration is not sufficient to correctly model the data. The optimal spin glass configuration is retrieved by an iterative minimization procedure, where chains are finally assimilated to fibers. As a result, all brain fibers appear as growing simultaneously until they merge with other fibers or reach the domain boundaries. In case of an ambiguity within a region like a crossing, the contribution of all neighboring fibers is used determine the correct neural pathway. This framework is tested on a MR phantom representing a 45° crossing and a real brain dataset. Notably, the framework was able to retrieve the triple crossing between the callosal fibers, the corticospinal tract and the arcuate fasciculus.

1 Introduction

The emergence of diffusion MRI for the past two decades enabled the in vivo study of anatomical connectivity via white matter tractography. Deterministic tractography algorithms reconstruct putative fascicles incrementally by following the main fiber directions as revealed by diffusion models (e.g, tensor or Q-Ball). However, such approach is prone to local errors in the estimate of the fiber directions (caused by noise and partial voluming) and may deviate from the true neural pathway. Probabilistic tractography [1,2,3] appeared as a way to handle the uncertainty of the fiber orientations by sampling each direction from a probability density function related to the diffusion model, and by repeating several thousands of times this technique from a seed point. While those methods was shown to be less noise-sensitive compared to their deterministic counterparts, their output is also very different: they return a connectivity map where each voxel is proportional to the probability of being connected to the seed. This raises the question of the statistical significance of those maps (which value is significant, which is not?). Moreover, further decomposition of fibers into morphological descriptors (e.g., length, curvature) for shape analysis becomes difficult. Recently, global alternatives to tractography were developed [4,5]. In those,

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 927–934, 2009.

the entire neural pathway is the parameter to be optimized, which elegantly adds robustness to deterministic tractography.

In this work, we propose a novel framework for global tractography which is a complete overhaul of the spin glass model introduced in [6].[1] The entire set of white matter fibers is parameterized by small segments called spins. Contrary to [6] (where spin positions and number are fixed), spins are allowed to move, rotate and duplicate. To control their degrees of freedom, spins are endowed with three potential energies: a diffusion, an interaction, and a generative potential. The diffusion potential attracts spins towards the main fiber directions, while the interaction potential encourages them to form longer chains of minimal curvature. The generative potential prevents a spin chain to end inside the domain by allowing the creation of new spins. The optimal spin configuration is finally retrieved by a global minimization procedure. As a result, all brain fibers appear as growing simultaneously until they merge with other fibers or reach the white matter boundaries. The advantages of such approach are fourfold: 1) it only relies on the two generally admitted priors that brain fibers have a low curvature and do not end inside white matter, 2) it does not require an estimation of the number nor directions of the fiber compartments in each voxel, 3) it automatically adapts the number of parameters (i.e., spins) to the data, and 4) it can be adapted to any type of diffusion model.

The rest of the paper is organized as follows. The framework for spin glass tractography (SGT) is exposed in Sec. 2. Experiments on real datasets are conducted in Sec. 3, before concluding in Sec. 4.

2 An Adaptive Spin Glass Model for Tractography

A spin s is an oriented particle defined by its position x and orientation $\overrightarrow{v}$ (of unit length). A spin glass S is an ensemble of N spins contained within a closed domain Ω: $S = \{s_i(x_i, \overrightarrow{v}_i)/x_i \in \Omega\}_{0 \leq i < N}$. A spin can be interpreted as a piece of fiber (point plus tangent), while Ω is the definition domain of the fibers.

Spins are endowed with three types of potential energy: a *diffusion* potential, an *interaction* potential and a *generative* potential. The diffusion potential acts as a non-stationary magnetic field attracting the spin orientations towards the main fiber directions. The interaction potential controls how spin associate with neighbors and embeds the prior that fibers have a low curvature. The generative potential authorizes the spontaneous generation of new spins to ensure that fibers do not end inside white matter. Note that the term "adaptive spin glass" is due to this last potential, which adapts the quantity of spins to the data. Fibers are finally retrieved by minimizing the sum of those three potentials over all spins in the glass. In the following, we first introduce each potential and present the algorithm for fiber reconstruction based on this spin glass model.

[1] We keep the terminology "spin glass" although we do not rigorously follow the concepts of the original spin glass model developed in statistical physics.

2.1 The Diffusion Potential

Within our framework, the minima of the diffusion potential E_d should coincide with the main fiber directions. The expression of this potential obviously depends on the diffusion model chosen to represent the data. However, we may follow this principle for a general definition. When the probability of finding a fiber in a direction is high, the potential should be low and conversely. Furthermore, when the probability is close to zero, the potential should tend towards infinity as it is very unlikely to find a fiber in that direction. Thus, E_d should be proportional to the log-likelihood of the fiber presence in a direction. One can write, for any model:

$$E_d(x, \overrightarrow{v}) = -\log\left(p(x, \overrightarrow{v})\right), \tag{1}$$

where $p(x, \overrightarrow{v})$ is the probability of existence of a fiber at location x and in direction $\overrightarrow{v}$. Assuming that the main tensor eigenvectors are good estimates of the fiber directions, we provide the following expression of p for a tensor field $D(x)$: $p(x, \overrightarrow{v}) = \overrightarrow{v}^{\top} D(x) \overrightarrow{v}$ ($D(x)$ should be normalized to ensure $\int_{S^2} p(x, \overrightarrow{v}) dv = 1$, S^2 being the unit sphere). Dedicated expressions can be derived for other models. In particular, the orientation distribution functions (ODF) [7] are appealing since they directly provide the probability p. Note that multiple maxima are possible: in this case, the diffusion potential will act as a multi-modal magnetic field and attract a spin towards one particular fiber direction depending on its initial orientation.

2.2 The Interaction Potential

The interaction potential E_{int} controls how spins interact with each others and encourages them to assemble into longer chains. Following [6], this potential embeds the prior that brain fibers have a low curvature. Basically, a spin orientation separates the spin glass into two sets: those which are behind it (backward spins) and those in front of it (forward spins). We define the set of forward (resp. backward) neighbors as $N^f(s) = \{s_i \in S / \|x_i - x\| < r \text{ and } \overrightarrow{v}.\overrightarrow{xx_i} > 0\}$ (resp. $N^b(s) = \{s_i \in S / \|x_i - x\| < r \text{ and } \overrightarrow{v}.\overrightarrow{xx_i} \leq 0\}$). r is the radius controlling the size of the neighborhood. A spin s associates with a unique backward $s^b(x^b, \overrightarrow{v}^b)$ and forward $s^f(x^f, \overrightarrow{v}^f)$ neighbor such that the interaction potential of their trajectory $s^b - s - s^f$ is minimal.

We propose to define the interaction potential between a spin and its backward and forward neighbors as:

$$\mathrm{E}_{int}(s, s^b, s^f) = -\frac{1}{5} \sum_{j=1}^{5} \log\left(\frac{\cos\alpha_j - \cos\alpha_{max}}{1 - \cos\alpha_{max}}\right), \tag{2}$$

where α_1 (resp. $\alpha_{2,3,4,5}$) is defined as the angle between $\overrightarrow{v}^b$ and $x - x^b$ (resp. $x - x^b$ and $\overrightarrow{v}$, $\overrightarrow{v}$ and $x^f - x$, $x^f - x$ and $\overrightarrow{v}^f$, $\overrightarrow{v}^b$ and $\overrightarrow{v}^f$) (see Fig. 1). α_{max} is the maximum angular deviation allowed and lie within the range $]0, \pi/2]$. Any angular difference greater than α_{max} will produce an infinite potential which

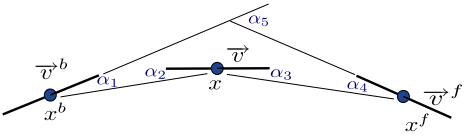


Fig. 1. Angular differences between a spin s and its backward (s^b) and forward (s^f) neighbors. A spin is represented by a blue circle (position), and a black line (orientation). The role of α_5 is to prevent the creation of a junction between two incompatible neighbors, which would result in an angular deviation greater than α_{max}.

forbids the spin association. Conversely, spins perfectly aligned will associate as their interaction potential is null. If a spin leads outside the glass, i.e., $x \pm \overrightarrow{v} \notin \Omega$, then no interaction potential is added for this half-neighborhood.

2.3 The Generative Potential

This last potential plays a particular role: it ensures that brain fibers do not end inside white matter, which can be translated by spin chains do not stop within the glass. There are situations where a spin cannot find a candidate neighbor to associate with, which causes a chain to end prematurely. This is the case when all neighbors have already been associated (we recall that a spin associates to a unique backward and forward neighbor), or when available neighbors lead to an infinite interaction potential (angular differences greater than α_{max}). In this case, the generative potential will become infinite, ordering the creation of a new spin to complete the broken chain. We formulate this potential as follows:

$$\begin{aligned} E_{gen}(s) &= \infty \quad \text{if } \forall (s^b, s^f) \in N^b(s) \times N^f(s), E_{int}(s, s^b, s^f) = \infty \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{3}$$

2.4 Global Potential and Algorithm Overview

The combination of the interaction (Eq. 1) and diffusion potentials (Eq. 2) has a competitive influence on the spin orientations and localizations. Hence, the optimal spin glass configuration corresponds to a trade-off between fidelity to the diffusion data and low curvature. The generative potential (Eq. 3) only affects the quantity of spins so that chains do not stop within the glass. The optimal spin configuration S^* is obtained when the sum over all spins of the three potentials is minimal, i.e.:

$$S^* = \min_{S=\{s_i\}} \sum_{i=1}^{N} E_d(s_i) + \lambda E_{int}(s_i, s_i^b, s_i^f) + E_{gen}(s_i), \tag{4}$$

where λ controls the trade-off between interaction and diffusion potentials.

The optimal configuration is retrieved via an alternate minimization procedure. For a given spin glass configuration, spin positions and orientations are optimized to minimize their diffusion and interaction potentials only. Once the

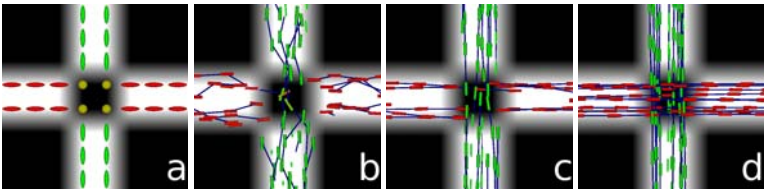


Fig. 2. Example of SGT of a synthetic fiber crossing. a) The tensor field. b) The initial spin glass: spins (represented by green and red cylinders) were randomly placed and aligned with the main tensor eigenvectors. The blue links represent the spin associations. c) The minimal energetic configuration of the spin glass in b). d) After convergence: spin chains have grown and merged to reconstruct the crossing area.

stable spin configuration is found, the generative potential is minimized by merging together compatible spin chains and by adding new spins at the extremities of remaining broken chains. During this phase, the contribution of all neighboring spin chains in regions of crossing fibers is used to determine the most likely pathways (the ones with the lowest possible curvature). Thus, there is no need for knowing the number of fiber compartments in each voxel, as those are automatically determined by the neighborhood. Finally, the process is reiterated until convergence (no more spin association and generation). Depending on the size of the dataset, from 10 to 100 iterations are necessary and tens to millions of spins are required. Example of SGT of a synthetic crossing is shown in Fig. 2 top row (20 iterations, 230 initial spins, 314 spins after convergence, $\lambda = 1.0$, $\alpha_{max} = 45°$, $r = 3mm$). In the next section, we evaluate the performance of the SGT algorithm on real datasets.

3 Experiments

We evaluated the spin glass algorithm on two datasets: a MR phantom consisting of a 45° crossing and a real brain dataset. For each dataset, we compared the output of the SGT algorithm to a tensor-based streamline tractography (TBT) (the tensorlines [8]) and a Q-Ball-based streamline tractography (QBT) [9] (generalization of classical streamline techniques to Q-Ball) algorithms. For tensor fitting, we used the Rician noise removal strategy developed in [10]. For Q-Ball estimation, we opted for the spherical harmonics decomposition of [9] which provides an analytical formulation of the ODF. Note that the diffusion potential of the SGT was derived from the ODF and not the tensor. Parameters of the spin glass algorithm are: $\lambda = 1$, $\alpha_{max} = 45°$ and $r = 3mm$. The same angular threshold was used for the TBT and QBT to obtain comparable results.

3.1 MR Phantom

The crossing phantom was elaborated with hydrophobic fibers strongly tightened with a medium and immersed in a solution of water doped with gadolinium [11]. Acquisitions were done on a 1.5T MRI scanner (4000 directions, b-value of

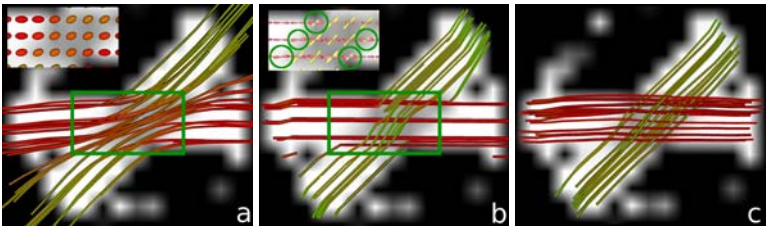


Fig. 3. SGT of the 45° crossing MR phantom. a) Result of TBT. Tensors contained within the green rectangle are displayed. b) Result of QBT. ODF contained within the green rectangle are displayed (those circled in greens had only one maxima out of two detected). c) Result of SGT. The same ODF were used for QBT and SGT.

2000s.mm^{-2}, image size: 32×32×3, resolution: 1×1×1.4mm). Results of tractography are shown in Fig. 3. Obviously, TBT was not able to recover a crossing of this low angular difference, since tensors are unable to model the two-fiber compartment inside the crossing area. QBT algorithm was able to recover part of the crossing but deviates in regions where the ODF peaks are not well defined (mainly due to noise and partial voluming). Indeed, maxima are generally detected by thresholding the ODF not to extract small noisy peaks. In our case, the ODF circled in green (Fig. 3 b) had only one maxima detected out of the two expected, which caused several erroneous pathways. SGT, however, was able to pass the crossing by using the neighborhood to infer the most likely pathways (Fig. 3 c).

3.2 Real Brain Dataset

The brain dataset was acquired on a 1.5T scanner with two protocols. First, a DT-MRI dataset of 41 directions and a b-value of 700s.mm^{-2} was acquired for TBT. Second, a HARDI dataset of 200 directions with a b-value of 3000s.mm^{-2} was acquired for QBT and SGT. Image size is 128×128×60, resolution is: 1.8 × 1.8 × 2mm. Results of trackings are presented in Fig. 4. For clarity reasons, only fibers passing through a sagittal slice of the corpus callosum (CC) are presented. Fibers are colored depending on the CC position they traverse (linear color gradient from the posterior in red to the anterior in blue of the CC). Nearly all fibers reconstructed by the TBT algorithm were redirected vertically because of the projection fibers (corona radiata) crossing the callosal fibers. QBT performed slightly better and fibers connecting the lateral part of the frontal and parietal lobes were found, showing that QBT is able to pass the corona radiata. However, the quality of the ODF was not good enough to allow QBT to recover the full set of association fibers crossing the CC. SGT was able to resolve this crossing area and exhibits many more interhemispheric connections. Moreover, the consistency of the coloring scheme indicates that the position of a fiber connecting the cortex is related to its anteroposterior position in the CC. SGT required 2 million spins to reconstruct the entire brain and three days of computations on a regular PC.

To further illustrate the performance of the SGT, we display in Fig. 5 the intersection between the CC, the corticospinal tract (CST), the arcuate

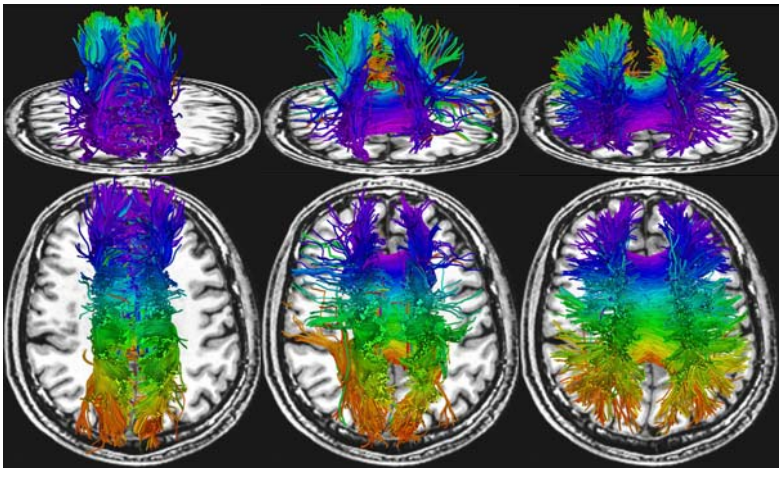


Fig. 4. Reconstruction of callosal fibers with three methods. Left: TBT. Middle: QBT. Right: SGT. All fibers of TBT are redirected vertically because of the surrounding corona radiata. QBT performed slightly better but missed a large part of the callosal fibers. SGT, by using the neighborhood to determine the most plausible pathways, was able to recover the myriad of fibers passing the corona radiata.

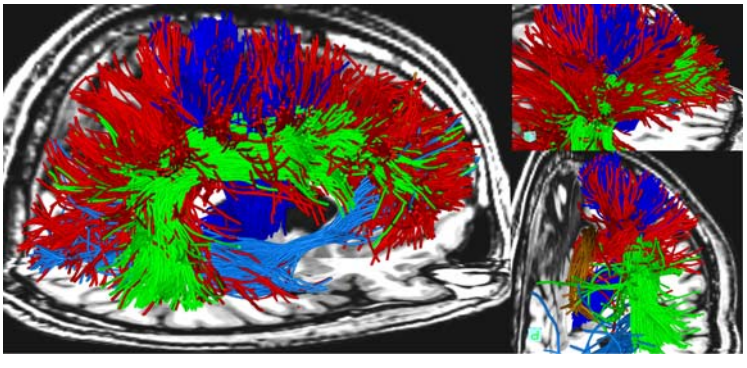


Fig. 5. Intersection between the CC (red), the CST (dark blue), the AF (green), the CB (orange) and the ILF (light blue) revealed by spin glass tractography. This region is one of the most complex crossing area accessible at this resolution of diffusion images.

fasciculus (AF), the cingulum bundle (CB), and the inferior longitudinal fasciculus (ILF). The triple-crossing region made by the CC, the CST and the AF was successfully reconstructed by the SGT algorithm.

4 Conclusion

In this work, we presented a new methodology for white matter fiber reconstruction inspired from a spin glass model. Spins are fiber elements endowed with a

diffusion, interaction and generative potentials. The combination of those three potentials allows the creation of long chains of spins of low curvature and fitting the diffusion data at best, which grow until they merge with other chains or reach the domain boundaries. In case of an ambiguity (like a crossing), the pathways leading to the lowest curvature are automatically chosen. Such approach has the advantage not to require a prior knowledge on the number of fiber compartments per voxel, and can be adapted to any type of diffusion model. Experiments on a synthetic MR phantom and a real brain dataset indicated that such a global approach is more successful than deterministic streamline tensor or q-ball-based methods to recover crossing fibers. Notably, we showed that spin glass tractography can successfully reconstruct the triple crossing between the corpus callosum, the corticospinal tract and the arcuate fasciculus, which is one of the most complex crossing area accessible at this level of image resolution.

References

1. Parker, G., Alexander, D.: Probabilistic anatomic connectivity derived from the microscopic persistent angular structure of cerebral tissue. Philosophical Transactions of the Royal Society B 360, 893–902 (2005)
2. Friman, O., Farneback, G., Westin, C.F.: A bayesian approach for stochastic white matter tractography. IEEE TMI 25(8), 965–978 (2006)
3. Jbabdi, S., Woolrich, M., Andersson, J., Behrens, T.: A bayesian framework for global tractography. NeuroImage 37, 116–129 (2007)
4. Zhang, F., Hancock, E.R., Goodlett, C., Gerig, G.: Probabilistic white matter fiber tracking using particle filtering and von Mises-Fisher sampling. MedIA 13(1), 5–18 (2009)
5. Kreher, B., Madeer, I., Kiselev, V.: Gibbs tracking: A novel approach for the reconstruction of neuronal pathways. MRM 60, 953–963 (2008)
6. Mangin, J.F., Poupon, C., Cointepas, Y., Riviére, D., Papadopoulos-Orfanos, D., Clark, C.A., Régis, J., Le Bihan, D.: A framework based on spin glass models for the inference of anatomical connectivity from diffusion-weighted MR data. NMR in Biomedicine 15, 481–492 (2002)
7. Tuch, D.: Q-ball imaging. MRM 52(6), 1358–1372 (2004)
8. Weinstein, D.M., Kindlmann, G.L., Lundberg, E.C.: Tensorlines: Advection-diffusion based propagation through diffusion tensor fields. In: IEEE Vis 1999, pp. 249–253 (1999)
9. Descoteaux, M., Deriche, R., Knösche, T., Anwander, A.: Deterministic and probabilistic tractography based on complex fibre orientation distribution TMI 28(2) (2009)
10. Fillard, P., Arsigny, V., Pennec, X., Ayache, N.: Clinical DT-MRI estimation, smoothing and fiber tracking with log-Euclidean metrics. TMI 26, 1472–1482 (2007)
11. Poupon, C., Rieul, B., Kezele, I., Perrin, M., Poupon, F., Mangin, J.F.: New diffusion phantoms dedicated to the study and validation of HARDI models. MRM 60, 1276–1283 (2008)

Tractography-Based Parcellation of the Cortex Using a Spatially-Informed Dimension Reduction of the Connectivity Matrix

Pauline Roca[1,2], Denis Rivière[1,2], Pamela Guevara[1,2], Cyril Poupon[1,2], and Jean-François Mangin[1,2]

[1] CEA Saclay, Neurospin/LNAO, Bât 145, 91191 Gif-sur-Yvette cedex, France
[2] Institut Fédératif de Recherche 49, Gif-sur-Yvette, France

Abstract. Determining cortical functional areas is an important goal for neurosciences and clinical neurosurgery. This paper presents a method for connectivity-based parcellation of the entire human cortical surface, exploiting the idea that each cortex region has a specific connection profile. The connectivity matrix of the cortex is computed using analytical Q-ball-based tractography. The parcellation is achieved independently for each subject and applied to the subset of the cortical surface endowed with enough connections to estimate safely a connectivity profile, namely the top of the cortical gyri. The key point of the method lies in a twofold reduction of the connectivity matrix dimension. First, parcellation amounts to iterating the clustering of Voronoï patches of the cortical surface into parcels endowed with homogeneous profiles. The parcels without intersection with the patch boundaries are selected for the final parcellation. Before clustering a patch, the complete profiles are collapsed into short profiles indicating connectivity with a set of putative cortical areas. These areas are supposed to correspond to the catchment basins of the watershed of the density of connection to the patch computed on the cortical surface. The results obtained for several brains are compared visually using a coordinate system.

1 Introduction

Characterization of functional areas and their relationships is a key to understand how brain works. Recent studies follow this goal and represent the brain as a graph, analyzing this network using graph theory [1,2,3]. The construction of this *human "connectome"* requires the definition of basic structural elements (nodes of the network), which can be chosen at different scales: level of single neurons (microscale), neuronal populations (mesoscale), or anatomically distinct brain regions (macroscale) [4]. In this context, a parcellation of the human brain into functional areas can be a solution to the definition of a first network at a macroscale. Following the idea that each brain module has a specific connectional fingerprint [5], the connectivity information based on diffusion-weighted imaging is an interesting candidate for defining the connectome nodes. A comparison with a functional connectivity study could be very useful for a validation [6,7].

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 935–942, 2009.

The main trouble to deal with when addressing connectivity-based parcellations from diffusion imaging is the huge dimension of the raw connectivity profiles. For instance, a standard mesh of the cortical surface is made up of more than 10 000 nodes. Segmenting this surface into parcels with homogeneous connectivity profiles amounts to performing clustering among vectors of dimension 10 000. Clustering is prone to the well-known curse of dimensionality. Therefore, reduction dimension seems mandatory. Several approaches have been proposed for this purpose. The first one is based on a segmentation of the brain to collapse the connectivity profiles: all the tracts reaching the same segment are summed up. This segmentation can be anatomical, for instance based on lobes or gyri [8,9,10], but the same idea could be applied with fMRI-based activation maps. In the papers cited above, the segmentation used for collapsing was based on individual data. This segmentation could also stem from the results of voxel-based group analysis or the results of invasive anatomical tracing performed in primates [11]. Such a priori knowledge is mapped into individual space using spatial normalization or sulcal landmarks. Finally, performing the clustering in a small area of the brain can overcome the curse of dimensionality. It was shown that Broca's area [12,13] and lateral premotor cortex [14] can be reproducibly parcellated from connectivity profiles towards the whole brain.

In this paper, the goal is to design a dimension reduction strategy efficient enough to allow reproducible parcellation of the whole cortical surface. We explore the possibility to reach this goal without using any information about the geometry of the cortical surface (sulco-gyral anatomy, coordinate systems). This is a challenging objective, but in our opinion, it has to be tackled to examine the extent to which one can get rid of hypotheses about the connectome structure that could bias the connectivity-based parcellation. The objective is all the more challenging since the parcellation is applied independently on each subject. Hence tractography errors cannot be averaged out by a group analysis. Our method is based on three key points that are described in the paper:

1. The algorithm is applied to the subset of the cortical surface endowed with enough connections to safely estimate a connectivity profile. This subset turns out to match the top of the cortical gyri. In this paper, we focus on long diffusion-based tracts that are supposed to convey more reproducible connectivity information than short U-fiber tracts.
2. The clustering is not performed on a whole brain basis but with an iterative patch by patch strategy.
3. The connectivity profiles are collapsed in an adaptive way for each patch to be parcellated: the segmentation for collapsing is based on the catchment basins of the watershed of the density of connection to the patch computed on the cortical surface.

An original visualization of the resulting parcellation has been developed. It consists in superimposing on each parcel a small replica of the whole brain indicating either the density of connection to the underlying parcel or the tracts terminating in this parcel. We compare the results obtained with four subjects using a coordinate system provided by BrainVISA based on the main sulci [15].

2 Materials and Method

2.1 Data and Pre-processings

The present study has been performed on four subjects of the NMR public database [16]. Diffusion weighted data were acquired with HARDI scheme, resulting in high-quality datasets based on 200 directions and a b value of 3 000 s/mm^2. Perfect alignment between diffusion data and a T1-weighted 1mm resolution image is achieved using a dedicated set of distortion corrections. The T1 pipeline of BrainVISA open software[1] is used to compute a mesh S of the cortical surface and a sulcus-based coordinate system used to compare subjects. The mesh is made up of two hemispheric spherical representations of the grey/white interface including about 40 000 vertices. The direct parcellation of the cortical surface capitalizes on the current knowledge of the columnar organization of the cortex: local populations of neurons form elementary processing units organized orthogonally to the cortex surface [4].

A deterministic tractography based on analytical Q-ball field [17] is performed using BrainVISA diffusion toolbox. The tracts are the trajectories of particles with inertia, which is leading to regularize curvature [18]. Tractography was initiated from two seeds in each voxel of the high resolution white matter mask computed from the T1-weighted acquisition, in both retrograde and orthograde directions, according to the maximal direction of the underlying ODF. Tracking is stopped either when the particle exits the propagation mask, when the angle between the two last moves exceeds 30°, or when the tract length exceeds 200 mm. Finally, tracts shorter than 20 mm were filtered out, leading to a set of about 1.5 millions tracts per subject.

Then histograms of tract lengths were computed for each subject (see Fig. 1) showing a reproducible two mode distribution. This study explores the connectivity information provided by the second mode defined by a 10 cm threshold. This ad hoc choice is based on the hypothesis that long range bundles are more reproducible across individuals than shorter U-fiber bundles. This threshold might at first glance look very high, but histograms show that about one third of the tracts are longer than 10cm.

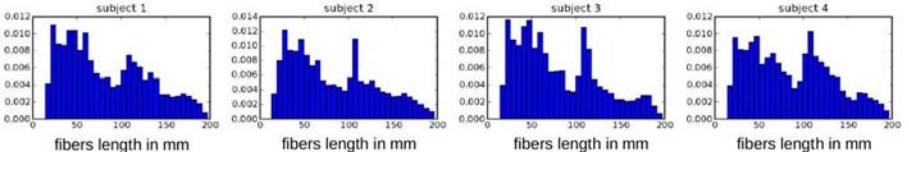


Fig. 1. Fibers length histograms

[1] http://brainvisa.info

2.2 Cortical Connection Matrix Construction

The cortex surface is divided into Voronoi cells centered around the surface mesh vertices. Then a raw connectivity matrix C^{raw} is computed by counting fibers connecting cells [19]. The raw matrix C^{raw} is then smoothed over the surface to account for a reasonable uncertainty on the tracking result. The smoothed matrix C used further results from the convolution of C^{raw} with a Gaussian. Gaussian weigths stem from geodesic distances computed using fast marching over triangulations [20]. The Gaussian standard deviation has been set to 5 mm.

2.3 Parcellation of the Cortex Mesh

The cortex surface S is parcellated according to an iterative algorithm summarized in Fig.2. The algorithm is applied to the subset of the cortical surface endowed with enough connections to safely estimate a connectivity profile. The density of connexions for each vertice is the L_1-norm of the corresponding row in C. A threshold on density provides the initial domain which turns out to correspond to the top of the gyri. In the following, we call resp. *cortex sub-mesh* connected component of the graph formed by the evoluting cortex triangulation S, *patch* Voronoï parcellation's segment, *parcel* final parcellation's component.

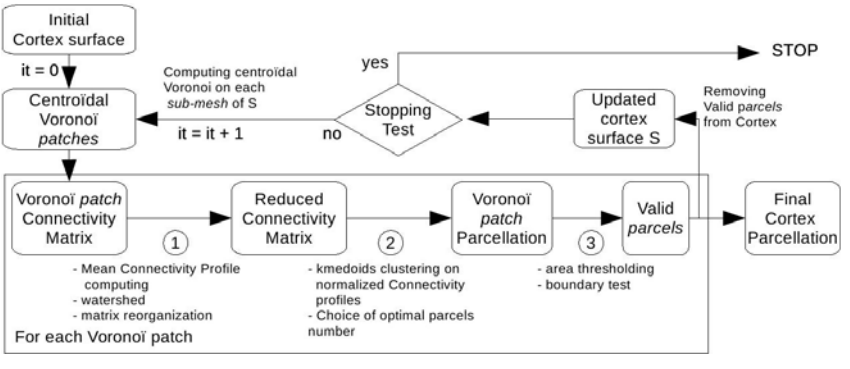


Fig. 2. A sketch of the parcellation Method

Each iteration begins with the random tessalation of the current domain into large *patches* of similar sizes. This splitting of the domain, aiming at reducing the complexity of the clustering problem, is performed *cortex sub-mesh* by *cortex sub-mesh*. Each *cortex sub-mesh* is split using the Centroïdal Voronoï tessalation algorithm [21]: a set of points is chosen randomly in the domain, a first Voronoï diagram is computed from these points, then the process is iterated using the Voronoï patch centroïds as seeds. We stop the process after five iterations. Assuming a value of 150 areas per hemisphere, each individual cortical area would occupy an average of 6 cm^2, with some much larger areas and some

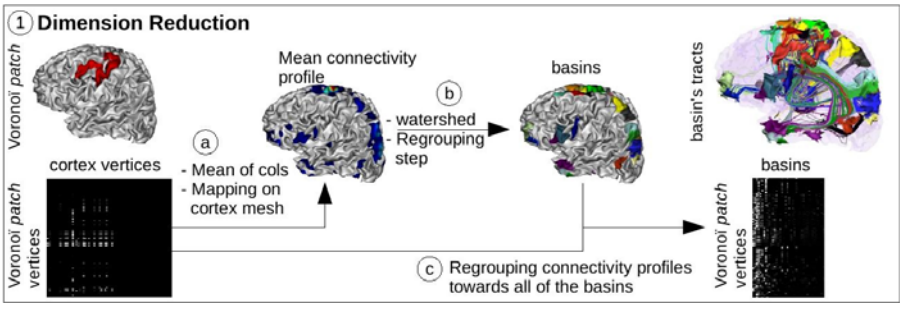


Fig. 3. Dimension Reduction Step 1. In the composite image on the right, the tract colors correspond to the color of the basin they connect with the Voronoi *patch*.

much smaller [22]. The number of random points used aims to split the domain into *patches* whose area correspond to ten times this average area of 6 cm^2.

Then each Voronoï *patch* is segmented into parcels with homogeneous connectivity profiles. The parcels larger than a minimum area and which do not intersect with the *patch* boundary are added to the final parcellation and removed from the domain during the next iterations. The algorithm stops as soon as one iteration did not add any *parcel* in the final result. The minimum area of a *parcel* is set to 2 cm^2, to account for the size variability of functional areas.

The clustering of the profiles of a Voronoï *patch* begins with an adaptive dimension reduction (Fig.3). The mean connectivity profile of the *patch* is computed and represented as a texture on the complete cortical surface. This texture is the density of connection to this *patch* (Fig.3.a). Then, a watershed is computed for this texture in order to split the cortical surface into catchment basins. The set of basins is pruned using a standard merging procedure based on depth and area. Each basin catches a set of tracts supposed to connect the *patch* with a meaningful brain area. The smallest basins are discarded in order to keep only 95% of the tract density (Fig.3.b) and the connectivity profiles are collapsed using the resulting set of basins (Fig.3.c). Finally, the reduced connectivity profiles are normalized (L_2-norm) and clustered using a kmedoids algorithm (PAM in R language [23]). The number of clusters K which gives the highest average silhouette width is considered as optimal [23].

3 Results

The method finds on average 140 *parcels* per hemiphere. The complexity of the data manipulated in this paper and the absence of gold standard prevents a straightforward evaluation of the results obtained for the four subjects. Hence we have decided to provide original visualizations that convey some insights about the level of reproducibility. A 3D object gathering all connectivity profiles is created. The parcellations are projected to an inflated cortex surface. For each *parcel*, a small cortex mesh textured with the normalized mean connectivity

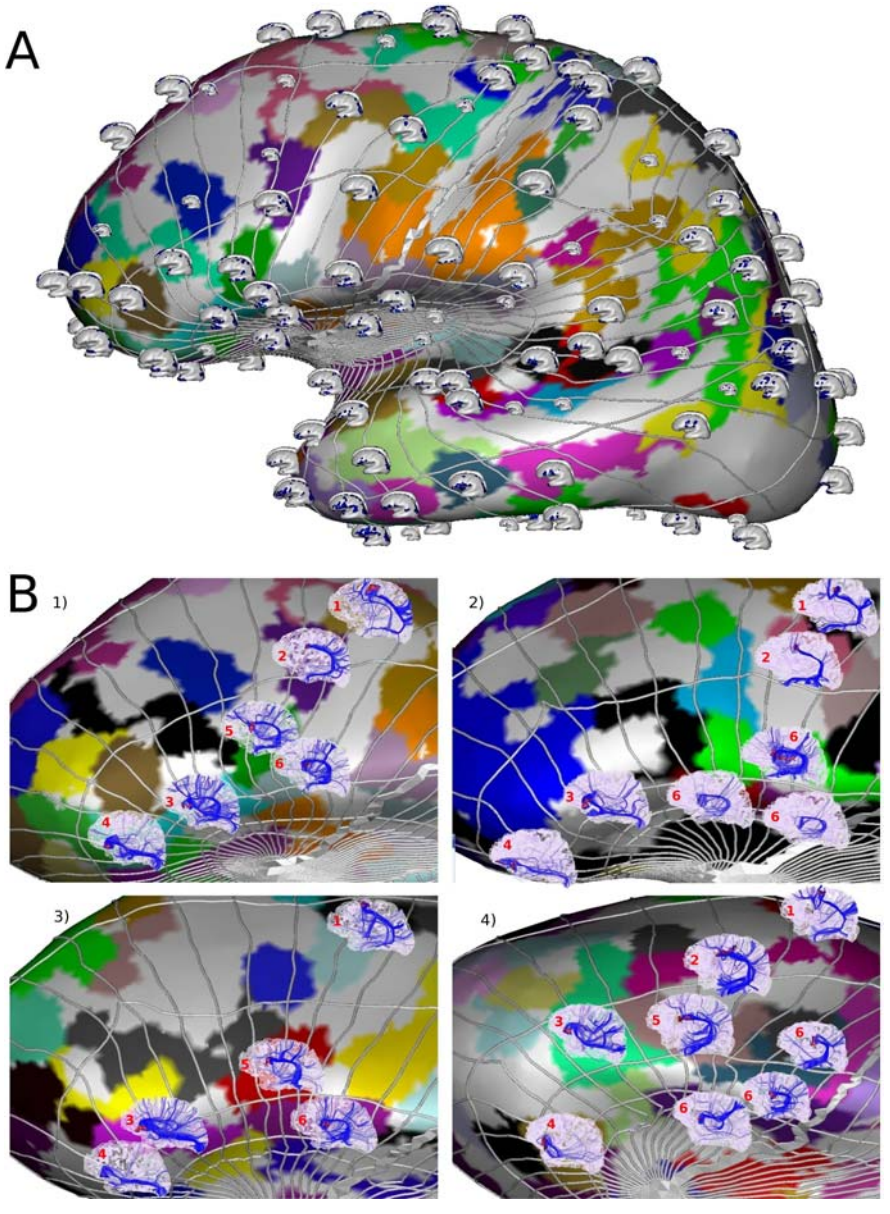


Fig. 4. A The parcels and their connectivity profiles for subject 1. The grid is a spherical coordinate system based on sulci. The thick line is the central sulcus. **B** Zoom on Parcellation around Broca's Area for the four subjects. A tract image is used to describe the profiles that show similarities across subjects.

profile of the *parcel* is put in the *parcel*'s centroïd. An additional structure representing meridians and parallels of a sulcus-based spherical coordinate system is superimposed [15] (see Fig.4.**A**).

While the coordinate system could allow the definition of quantitative evaluations of the reproducibility of the parcellation, we prefer to rely on visualization. The localization power of sulci relative to fiber bundles, is indeed a subject of research. The 3D scenes of profiles can be rotated in 3D in order to explore similarities across subjects. Additional windows can be opened to visualize the tracts: clicking on a *parcel* selects the connected tracts. We use this second functionality to create composite images of the 4 subjects centered around Broca's area. 3D images of the tracts were superimposed on the *parcels* whose profiles could be visually matched across subjects (see Fig.4.**B**). Tract images provide more information than profile images.

4 Discussion

Achieving a parcellation of the human brain cortex is fraught with difficulties, especially in the context of this study. A connectivity-based parcellation without anatomical a priori and performed independently for each subject is very ambitious. This study is still largely exploratory, relying on several parameters such as tracking method, fiber length threshold, clustering algorithm or optimal cluster's number, whose influence should be studied. Although this paper does not reach clearly reproducible results, the original visualization presented in this work leads to a better understanding of connectivity data and could also be used for studies of functional connectivity or any other vertex-based correlation measure.

Diffusion-based tractography is a rapidly developing field. The new generation of 1mm resolution acquisition that can now be achieved with highly parallel imaging and high fields could qualitatively change the individual parcellation obtained by our method. Furthermore, using the most advanced ODF deconvolution tools or probabilistic tracking could also improve our results. Nevertheless, the lack of reproducibility highlighted in this paper will lead us to address the parcellation at the level of the group of subjects. Surface-based coordinate system can be used to design such group studies. Other improvements will stem from using a segmentation of the deep structures (thalamus, basal ganglia, etc.) to increase the information coded in the connectivity profiles. Finally, a threshold free approach will be interesting for tuning the contributions of each tract to the profile according to its length.

References

1. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Neuroscience 10 (2009)
2. Iturria-Medina, Y., Canales-Rodriguez, E., Melie-Garcia, L.: Characterizing brain anatomical connections using diffusion weighted mri and graph theory. NeuroImage 36, 645–660 (2007)

3. Hagmann, P., Cammoun, L., Gigandet, X., et al.: Mapping the structural core of human cerebral cortex. PLOS Computational Biology 6(7), 1479–1493 (2008)
4. Sporns, O., Tononi, G., Kötter, R.: The human connectome: A structural description of the human brain. PLOS Comp. Biology 1(4), 245–251 (2005)
5. Passingham, R., Stephan, K., Kötter, R.: The anatomical basis of functional localization in the cortex. Nature 3, 606–616 (2002)
6. Honey, C.J., Sporns, O., et al.: Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. 106(6), 2035–2040 (2009)
7. Skudlarski, P., Jagannathan, K., Calhoun, V., et al.: Measuring brain connectivity: Diffusion tensor imaging validates resting state temporal correlations. NeuroImage 43(3), 554–561 (2008)
8. Behrens, T., Johansen-Berg, H., Woolrich, M.W., et al.: Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nature Neuroscience 6(7), 750–757 (2003)
9. Perrin, M., Cointepas, Y., Cachia, A.: Connectivity-based parcellation of the cortical mantle using q-ball diffusion imaging. Int. J. Biomed. Imaging (2008)
10. Guevara, P., Perrin, M., Cathier, P.: et al.: Connectivity-based parcellation of the cortical surface using Q-ball imaging. In: 5th Proc. IEEE ISBI, Paris, France, pp. 903–906 (2008)
11. Rushworth, M., Behrens, T., Johansen-Berg, H.: Connections patterns distinguish 3 regions of human parietal cortex. Cerebral Cortex 16, 1418–1430 (2005)
12. Anwander, A., Tittgemeyer, M., von Cramon, D., et al.: Connectivity-based parcellation of brocas area. Cerebral Cortex 17, 816–825 (2007)
13. Klein, J., Behrens, T., Robson, M., et al.: Connectivity-based parcellation of human cortex using diffusion mri: Establishing reproducibility, validity and observer independence in ba 44/45 and sma/pre-sma. NeuroImage 34, 204–211 (2007)
14. Tomassini, V.: Relating connectional architecture to grey matter function in the human lateral premotor cortex using functional and diffusion imaging. HBM (2008)
15. Clouchoux, C., Coulon, O., Rivière, D.: Anatomically constrained surface parameterization for cortical localization. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 344–351. Springer, Heidelberg (2005)
16. Poupon, C., Poupon, F., Allirol, L.: NMR: a free database dedicated to the anatomofunctional study of the human brain connectivity. In: HBM (2006)
17. Descoteaux, M., Deriche, R., Knösche, T.R., et al.: Deterministic and probabilistic tractography based on complex fibre orientation distributions. IEEE Transactions on Medical Imaging 28, 269–286 (2009)
18. Perrin, M., Poupon, C., Cointepas, Y.: Fiber tracking in q-ball fields using regularized particle trajectories. Inf. Process Med. Imaging 19, 52–63 (2005)
19. Cathier, P., Mangin, J.-F.: Registration of cortical connectivity matrices. In: Proc. MMBIA 2006, New York, USA (2006)
20. Kimmel, R., Sethian, J.A.: Computing geodesic paths on manifolds. Proc. Natl. Acad. Sci. 95, 8431–8435 (1998)
21. Du, Q., Faber, V., Gunzburger, M.: Centroidal voronoi tessellations: Applications and algorithms. Society for Industrial and Applied Mathematics Review 41(4), 637–676 (1999)
22. Van Essen, D.C., Dierker, D.L.: Surface-based and probabilistic atlases of primate cerebral cortex. Neuron 56, 209–225 (2007)
23. Kaufmann, L., Rousseeuw, P.: Finding groups in data: an introduction to cluster analysis. Wiley Interscience, Hoboken (1990)

Belief Propagation Based Segmentation of White Matter Tracts in DTI

Pierre-Louis Bazin, John Bogovic, Daniel Reich, Jerry L. Prince, and Dzung L. Pham

Johns Hopkins University, Baltimore, USA

Abstract. This paper presents a belief propagation approach to the segmentation of the major white matter tracts in diffusion tensor images of the human brain. Unlike tractography methods that sample multiple fibers to be bundled together, we define a Markov field directly on the diffusion tensors to separate the main fiber tracts at the voxel level. A prior model of shape and direction guides a full segmentation of the brain into known fiber tracts; additional, unspecified fibers; and isotropic regions. The method is evaluated on various data sets from an atlasing project, healthy subjects, and multiple sclerosis patients.

1 Introduction

Diffusion-weighted imaging (DWI) has become a major tool for the study of connectivity in the living human brain in health and disease [1,2]. Although DWI tractography methods provide useful information about the connectivity between remote regions of the brain [3], the segmentation of fiber bundles or tracts that correspond to known anatomical atlases requires additional grouping, trimming, and labeling, which usually requires manual assistance [4,5]. The development of automatic solutions for white matter tract segmentation is a challenging problem for several reasons. For example, since the reliability of fiber tracking varies with imaging resolution, noise, and patient orientation [6], it is difficult to consistently recover the same fiber tracts in repeated observations of the same individual. As well, lesions in the white matter caused by disease or aging can impair fiber tracking, resulting in poor definition of the tracts.

As an alternative to the reconstruction of fiber samples, level set [7] and non-parametric fuzzy classification methods [8] have been investigated for the segmentation of a given bundle from a set of initial regions of interest (ROIs). These methods focus on a single tract, and often require a careful initialization in order to succeed. Deformable atlas registration techniques segment the white matter into homogeneous regions rather than tracts [9,10], but the relationship between these regions and the tracts is unclear. Probabilistic connectivity methods evaluate the connection strength between regions or voxels [11,12,13], but do not explicitly recover fiber tracts.

In this work, we propose a Markov Random Field (MRF) to model the diffusion properties and a belief propagation (BP) technique to estimate the most likely tracts at every voxel. Shape and direction priors are used to identify the tracts, and the MRF connects regions along their diffusion direction. With this approach, we can automatically and simultaneously segment multiple tracts with known anatomical and functional meaning, reduce the variability of the estimated tracts without manual initialization or post-processing, and handle robustly the presence of white matter lesions.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 943–950, 2009.

2 Methods

2.1 Problem Description and Notations

The goal of this work is to extract from this data set several of the major fiber tracts of the human brain: anterior thalamic radiation (ATR), cortico-spinal tract (CST), body of the corpus callosum (CCB), forceps of the corpus callosum (CCF), cingulum (CG), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus (SLF), inferior fronto-occipital fasciculus (IFO), uncinate fasciculus (UNC). These tracts are first defined using a probabilistic atlas of shape and direction co-registered to the studied images, and then refined by estimating the most likely diffusion pathways from the FA and diffusion directions by belief propagation.

Starting from a series of diffusion-weighted images, we first obtain a set of diffusion eigenvectors and associated eigenvalues $\{\boldsymbol{v}_j^n, \lambda_j^n\}_{1,N}$ with the standard linear reconstruction method [2], as well as a fractional anisotropy (FA) map a_j for each voxel j of the image. Extra-cranial tissues are removed with a semi-automatic method [14] applied to the mean diffusivity image.

When manipulating direction vectors $\boldsymbol{v}_j$, we denote $\bar{v}_j$ the unsigned orientation axis without direction, i.e., $+\boldsymbol{v}_j$ and $-\boldsymbol{v}_j$ correspond to the same orientation $\bar{v}_j$. We use the following definitions for the inner product and addition of these orientations:

$$\bar{v}_i \cdot \bar{v}_j = |\boldsymbol{v}_i \cdot \boldsymbol{v}_j|, \qquad \bar{v}_i + \bar{v}_j = \boldsymbol{v}_i + \text{sign}(\boldsymbol{v}_i \cdot \boldsymbol{v}_j)\boldsymbol{v}_j. \tag{1}$$

Finally, the eigenvalues are ranked and normalized so that the largest eigenvalue $\lambda_j^1 = 1$.

2.2 Diffusion Based Tract Gain Functions

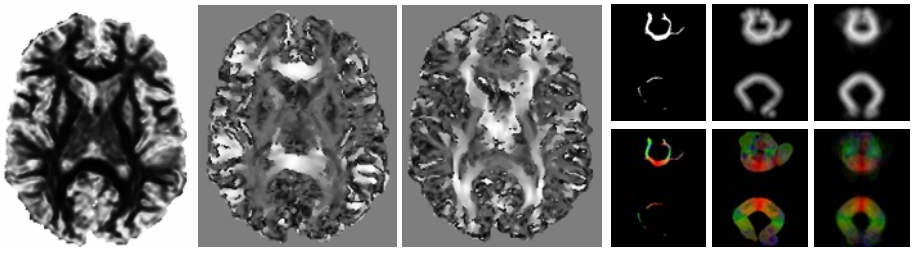
The central question for this approach is how to adequately translate our knowledge of fiber tracts observed in DTI into a MRF model of the tracts, represented by gain functions $L_{ij}(x_i, x_j)$, which we define to be a large positive value if there is evidence that the state x_i is likely conditioned on the state x_j, a large negative value if the data goes against this hypothesis, and close to zero if there is no information for or against it.

Several well-known fiber tracts like the optic radiation or even the corpus callosum were not included in the original, tractography-based atlas because it is difficult to define them consistently based on tractography and ROIs [6]. In addition, many smaller tracts between neighboring regions are present and distinct from these tracts. To provide a complete parcellation, we add two labels to the nine fiber tracts from the atlas: isotropic regions or background (BG), and other fiber tracts (OFT).

Background Model. We assume the background is composed of regions of very low FA, with the following gain function:

$$L_i^B = 2\exp(-\frac{a_i}{a_0}) - 1 \tag{2}$$

with a_0 a parameter representing the expected transition value for FA ($a_0 = 0.1$ in our experiments, see Fig. 1a).



a b c

Fig. 1. Gain and atlas functions: a) background gain L_i^B, b) diffusion gains L^P for the X and Y directions, c) shape and direction atlas building steps for CCF (left: original delineation, middle: extrapolated image, right: average of seven images)

Diffusion Model. The evidence for fiber tracts comes from the diffusion tensors: if two tensors are aligned they likely correspond to the same tract. Note that we cannot differentiate between the type of tracts based on the diffusion direction alone, without considering some global models of the tracts. However, we can model the similarity between tensors as follows:

$$L^P_{j|i} = \text{maxabs}_{n,m} \lambda_i^n \lambda_j^m \exp(-\frac{1 - \max(\bar{v}_i^n \cdot \bar{d}_{ij}, \bar{v}_j^m \cdot \bar{d}_{ij})}{v_1}) \left(2\exp(-\frac{1 - \bar{v}_i^n \cdot \bar{v}_j^m}{v_1}) - 1\right) \tag{3}$$

with $\bar{d}_{ij}$ the direction between voxels i and j, and v_1 a constant representing the average expected deviation in angle between similar directions. $\text{maxabs}_n g(n)$ selects for the function $g(n)$ with maximum absolute value.

This gain function is positive when the diffusion directions from i to j and from j to i are aligned with the path from i to j. If both diffusion directions are orthogonal to that path, we cannot assume that they are related even if they are aligned: many fiber tracts have "kissing" fibers that follow the same direction before diverging. In such case, the gain goes to zero in order to model the uncertainty. When the diffusion directions are orthogonal and one of them is aligned with the path, then it is clear that both points cannot be part of the same tract, which translates into a large negative gain value (see Fig. 1b).

Shape and Direction Prior. To provide a labeling for the tracts, we use an atlas of shape and direction that gives for each voxel i and tract label l a probability p_i^l of existence, and a probable diffusion orientation $\bar{d}_i^l$, with $|\bar{d}_i^l|$ a function of the orientation variability in the atlas.

Our atlas is based on the fiber tract atlas of Mori and Wakana [4,6]. We obtained the individual tensor images and fiber tract delineations used in building this atlas, so we could complement the atlas with tracts that were not originally delineated because of their lower reproducibility [6]. For this work, we added our delineation of the body of the corpus callosum in seven atlas images (see Fig. 2). The remaining, unspecified fibers were segmented as the regions of FA above $a_0 = 0.1$.

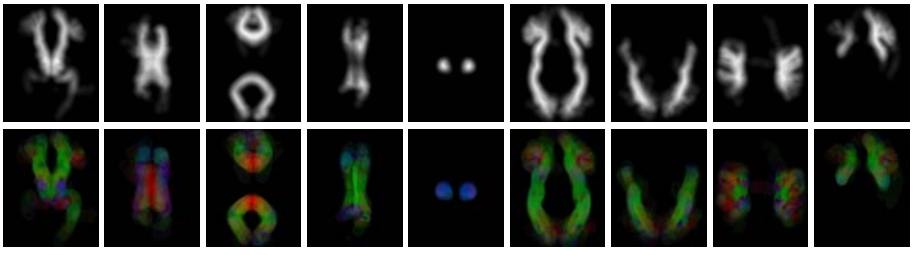


Fig. 2. Shape and direction atlas for the delineated fiber tracts on selected axial slices: shape prior p^l (top) and direction prior $\bar{d}^l$ (bottom) for each tract

To build the shape and direction atlas, each of the delineated fibers is smoothed over neighborhood of 10mm and its principal orientation is extrapolated as follows:

$$\bar{v}_i^t = \frac{\sum_j s_j^t \bar{v}_j^t}{\sum_j s_j^t} \quad \forall j \in N_i \text{ s.t. } s_j^t > s_i^t, \tag{4}$$

where $\bar{v}_i^t$ is the principal direction at voxel i in image t and s_i^t the smoothed delineation on image t for the considered fiber. The shape and orientation are then averaged from the seven images as shown in Fig. 1-c. The 10mm neighborhood was chosen empirically to ensure that tracts from the image to segment are fully included in the neighborhood, making the atlas more generic, even with low numbers of atlas images. As the segmentation is driven simultaneously by the atlas and the diffusion data, it is not necessary that the atlas matches accurately the images to segment.

From the atlas, we define the shape prior gain as follows:

$$L_i^S(x_i = l) = p_i^l - \max_{m \neq l} p_i^m \tag{5}$$

giving positive values where the prior for l is higher than any other, negative values where is it dominated by another prior, and zero where two or more tracts compete. The direction prior gain is similar to the diffusion gain:

$$L_i^D(x_i = l) = p_i^l |\bar{d}_i^l| \text{maxabs}_n \lambda_i^n \left(\exp(-\frac{|\bar{d}_i^l| - \bar{d}_i^l \cdot \bar{v}_i^n}{|\bar{d}_i^l| v_1}) - \exp(-\frac{\bar{d}_i^l \cdot \bar{v}_i^n}{|\bar{d}_i^l| v_0}) \right) \tag{6}$$

with v_0 an angular constant. The gain is positive when the data is well aligned with the prior, negative when it is orthogonal, and goes toward zero in the uncertain area between v_0 and v_1 or when the shape prior is low, i.e., where the orientation is likely to be unknown. For the unspecified fibers, we set $L_i^D(\text{OFT}) = 0$.

Gain Function. The complete gain function is built from these separate terms for all the different cases as follows:

$$\begin{aligned} L_{i|j}(\text{BG}|\text{BG}) &= L_i^B + \alpha L_i^S(\text{BG}) + \beta \\ L_{i|j}(\text{BG}|l) &= L_i^B + \alpha L_i^S(\text{BG}) \\ L_{i|j}(l|\text{BG}) &= -L_i^B + \alpha L_i^S(l) + \gamma L_i^D(l) \\ L_{i|j}(l|l) &= L_{i|j}^P \, p_{bj}(l) + \alpha L_i^S(l) + \beta + \gamma L_i^D(l) \end{aligned} \tag{7}$$

$$L_{i|j}(l|m) = -\max(L^P_{i|j}, 0)\, p_{bj}(l) + \alpha L^S_i(l) + \gamma L^D_i(l)$$

where $p_{bj}(l) = \frac{\exp b_j(l)}{\sum_m \exp b_j(m)}$ is the current estimate of the posterior probability for label l, and α, β, γ three parameters that modulate the importance of shape priors (α), smoothing (β), and direction priors (γ). Because isotropic regions do not include a direction prior, the relative value of γ with regard to α, β influences the global amount of recovered tracts. In all our experiments, we set $\alpha = 0.1$, $\beta = 0.2$ and $\gamma = 0.5$.

2.3 Belief Propagation Algorithm

Belief propagation [15] is the method of choice for maximizing functions that can be decomposed into a sum or product of pairwise interactions $E = \sum_{i,j} L_{i|j}(l|m)$ as above. In this work, we use the tree-reweighted belief propagation (TRBP) variant of Kolmogorov, which offers convergence guarantees not present in the original max-product algorithm [16].

First we register the atlas to the image to segment with a multi-scale gradient descent method maximizing $E_{R0} = \sum_j \sum_l ||a_j p^l_{T(j)}||^2$, where a_j is the computed FA, and T a rigid transform. The direction atlas is rotated accordingly. Then, the algorithm alternates belief propagation with a refinement of the atlas registration:

1. compute the beliefs: $b_i(x_i) \leftarrow \sum_{j \in N_i} m_{ji}(x_i)$,
2. estimate the gain function $L_{i|j}$ for all labels,
3. update the messages: $m_{ij}(x_j) \leftarrow \max_{x_i} \left(\frac{|N_i|}{2} L_{ij}(x_i, x_j) + \frac{2}{|N_i|} b_i(x_i) - m_{ji}(x_i) \right)$
4. refine the registration by maximizing $E_R = \sum_j \sum_l ||p_{bj}(l) p^l_{T(j)}||^2$

These steps are repeated until the computed beliefs are stable, only about 10-20 iterations in practice thanks to the efficient TRBP propagation method. A $181x217x181$ voxel image (1mm cubic resolution) is processed in less than two hours, and a more classical $256x256x60$ voxel image takes between 45 minutes and an hour on a modern workstation with 6GB of available memory.

3 Experiments

The algorithm is evaluated in several experiments that test its accuracy and reproducibility in real DTI images of the human brain in health and disease, as well as its use in complement to more classical tractography methods. First, we evaluate the algorithm on a set of ten images from the original atlas [4], distinct from the seven images used in building the shape and direction priors ("atlas images" experiment). In clinical practice, two or more DWI acquisitions are often used to reconstruct the DTI tensors in order to mitigate the noise. We reconstructed two separate tensor images from the DWI acquisitions for a set of five healthy subjects. These images have low SNR but describe the exact same anatomy in the same orientation ("healthy repeats" experiment). We also tested the method on a set of seven pairs of DTI acquisitions from multiple sclerosis (MS) patients with extensive white matter lesions, imaged twice at an interval of about six month ("MS repeats" experiment). Finally, we investigate the use of this voxel-based

Table 1. Overlap and distance results for the experiments: average values (standard deviation)

		Structures								
		ATR	CCB	CCF	CG	CST	IFO	ILF	SLF	UNC
Atlas images	Inclusion ratio	0.60 (0.10)	n/a n/a	0.67 (0.06)	0.45 (0.09)	0.73 (0.08)	0.50 (0.08)	0.36 (0.09)	0.49 (0.07)	0.32 (0.10)
Atlas images	Surface distance	3.90 (1.25)	n/a n/a	2.40 (0.25)	3.99 (2.00)	3.49 (1.09)	4.26 (3.53)	3.31 (0.48)	3.70 (0.89)	4.12 (1.31)
Healthy repeats	Dice overlap	0.64 (0.10)	0.70 (0.17)	0.61 (0.11)	0.48 (0.13)	0.73 (0.09)	0.48 (0.19)	0.39 (0.20)	0.54 (0.12)	0.26 (0.19)
Healthy repeats	Surface distance	0.87 (0.29)	0.83 (0.78)	0.88 (0.37)	1.38 (0.94)	0.92 (0.42)	1.24 (0.76)	1.27 (0.64)	1.03 (0.49)	2.30 (1.42)
MS repeats	Dice overlap	0.57 (0.11)	0.55 (0.15)	0.56 (0.07)	0.35 (0.18)	0.65 (0.12)	0.42 (0.13)	0.33 (0.10)	0.39 (0.12)	0.22 (0.13)
MS repeats	Surface distance	1.50 (0.62)	1.51 (0.91)	1.47 (0.47)	3.01 (1.87)	1.51 (0.84)	1.84 (0.95)	2.32 (1.15)	2.03 (1.10)	2.66 (1.36)

method in conjunction with fiber tractography. Tractography does provide a much finer, sub-voxel representation of connectivity, but automatically labeling regions and bundles of fibers is challenging [17,18]. Our method provides a probabilistic labeling of the entire space, thus we can classify each fiber by its mean belief value.

In DTI segmentation, validation is a challenging issue as there is no accepted gold standard or ground truth beyond simplistic simulations. Even the carefully edited fiber tracts of [4] have limitations, and often portions of the tracts are missing if the underlying fibers are interrupted or stray into a neighboring bundle. Furthermore, the typically elongated shape and small volume of fibers make the classical measures of overlap very sensitive to differences.

For the atlas images, we measured the amount of the original delineation (D) included in the segmentation (S): $I = \frac{D \cap S}{D}$, while we used the Dice overlap $D = \frac{2S_1 \cup S_2}{S_1 + S_2}$ in the repeats. Average surface distances were computed in all cases, see Table 1. In this context, our experiments still demonstrate that our segmentations correlate well with fiber delineations, and more importantly that they are reproducible in successive acquisitions of clinical quality.

As can be seen on Fig. 3-a, the tractography-based delineations of the atlas were very conservative, and our proposed segmentation is overall more inclusive. For the largest, most distinct tracts (ATR, CC, CST) the segmentation includes most of the original delineations, but the smaller (CG, UNC) and heavily overlapping tracts (IFO, ILF, SLF) are more variable, due to the small number of voxels in the delineation. The repeat experiments show fairly high overlap and very low surface distance for most tracts, and the differences in the resulting segmentations appear mostly to be related to noise, see Fig. 3-b. The method succeeds in the presence of lesions, however with slightly worse overlap than in the healthy cases. This is likely reflecting the added variability due to different head position, different geometric distortions and registration errors. The presence of lesions only impairs the segmentation locally, leading to thinner or missing tracts in some of the diseased regions, but the healthy parts of the tracts remain unaffected, see Fig. 3-c. When tractography is reliable, our method offers an efficient way to cluster the fibers into compact bundles, mostly free of straying fibers as shown in Fig. 3-d.

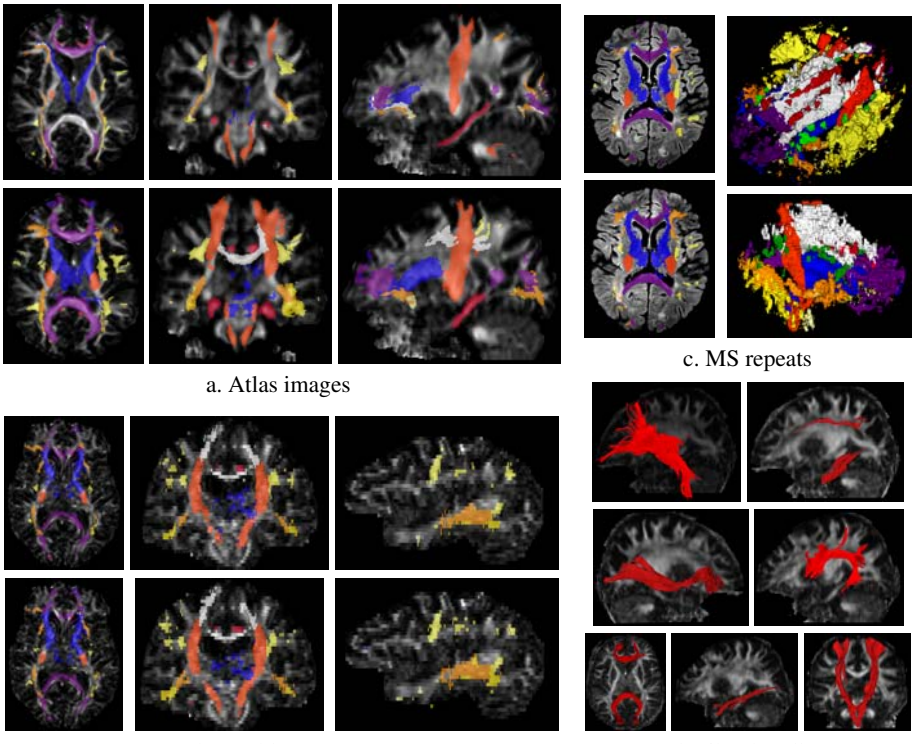


Fig. 3. Experiments: a) Atlas images: automatic segmentation (top) compared to tractography-based delineation (bottom), b) Healthy repeats: segmentation for two successive acquisitions, c) MS repeats: segmentations of two separate acquisitions, after coregistration (left) and 3D rendering of the reconstructed fibers intersecting estimated lesions (in green) in the same subject, d) fibers clustered by our segmentation in an atlas image

4 Conclusion

In this paper, we propose a new approach to the segmentation of white matter tracts in DTI. By combining shape and direction priors with a belief propagation method, the algorithm recovers nine of the major fiber tracts in the human brain automatically and simultaneously. Several real data experiments indicate that the method is consistent with manual delineations based on tractography, and can handle clinical quality images with noise and even lesions in a reproducible way. These properties are important for clinical applications, where the interaction of white matter lesions or tumors with the main fiber tracts can inform the diagnosis and treatment for individual patients and provide more insight on the systems targeted by the disease.

Acknowledgments. We thank Dr. Susumu Mori for providing us the original tensor images and tract delineations for the atlas used in this work [4]. This work was supported in part by the NIH/NIDA grant 1K25DA025356 and NIH/NINDS grant 1R01NS056307.

References

1. Bihan, D.L., Mangin, J., Poupon, C., Clark, C., Pappata, S., Molko, N., Chabriat, H.: Diffusion tensor imaging: concepts and applications. J. MR Imaging 13(4), 534–546 (2001)
2. Basser, P.J., Jones, D.K.: Diffusion-tensor MRI: theory, experimental design and data analysis - a technical review. NMR Biomed. 15(7-8), 456–467 (2002)
3. Mori, S., van Zijl, P.C.: Fiber tracking: principles and strategies - a technical review. NMR Biomed. 15(7-8), 468–480 (2002)
4. Mori, S., Wakana, S., Nagae-Poetscher, L.M., van Zijl, P.C.: MRI Atlas of Human White Matter. Elsevier, Amsterdam (2005)
5. Lawes, I., Barrick, T., Murugam, V., Spierings, N., Evans, D., Song, M., Clark, C.: Atlas-based segmentation of white matter tracts of the human brain using diffusion tensor tractography and comparison with classical dissection. Neuroimage 39, 62–79 (2008)
6. Wakana, S., Caprihan, A., Panzenboeck, M., Fallon, J., Perry, M., Gollub, R., Hua, K., Zhang, J., Dubey, P., Blitz, A., van Zijl, P., Mori, S.: Reproducibility of quantitative tractography methods applied to cerebral white matter. NeuroImage 36, 630–644 (2007)
7. Lenglet, C., Rousson, M., Deriche, R.: DTI segmentation by statistical surface evolution. IEEE Trans. Medical Imaging 25(6), 685–700 (2006)
8. Awate, S.P., Gee, J.C.: A fuzzy, nonparametric segmentation framework for DTI and MRI analysis. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 296–307. Springer, Heidelberg (2007)
9. Mori, S., Oishi, K., Jiang, H., Jiang, L., Li, X., Akhter, K., Hua, K., Faria, A.V., Mahmood, A., Woods, R., Toga, A.W., Pike, G.B., Neto, P.R., Evans, A., Zhang, J., Huang, H., Miller, M.I., van Zijl, P., Mazziottad, J.: Stereotaxic white matter atlas based on diffusion tensor imaging in an ICBM template. Neuroimage 40, 570–582 (2008)
10. Yang, J., Shen, D., Davatzikos, C., Verma, R.: Diffusion tensor image registration using tensor geometry and orientation features. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 905–913. Springer, Heidelberg (2008)
11. Parker, G.J.M., Alexander, D.C.: Probabilistic monte carlo based mapping of cerebral connections utilising whole-brain crossing fibre information. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 684–695. Springer, Heidelberg (2003)
12. Fletcher, P.T., Tao, R., Jeong, W.-K., Whitaker, R.T.: A volumetric approach to quantifying region-to-region white matter connectivity in diffusion tensor MRI. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, pp. 346–358. Springer, Heidelberg (2007)
13. Yendiki, A., Stevens, A., Jbabdi, S., Augustinack, J., Salat, D., Zollei, L., Behrens, T., Fischl, B.: Probabilistic diffusion tractography with spatial priors. In: Proc. MICCAI wksp. Computational Diffusion MRI (2008)
14. Goldszal, A., Davatzikos, C., Pham, D., Yan, M., Bryan, R., Resnick, S.: An image processing system for qualitative and quantitative volumetric analysis of brain images. J. Computer Assisted Tomography 22(5), 827–837 (1998)
15. Pearl, J.: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann, San Francisco (1988)
16. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE Trans. Pattern Analysis and Machine Intelligence 28(10) (2006)
17. Zhang, W., Olivi, A., Hertig, S.J., van Zijl, P., Mori, S.: Automated fiber tracking of human brain white matter using diffusion tensor imaging. Neuroimage 42, 771–777 (2008)
18. Mayer, A., Zimmerman-Moreno, G., Greenspan, H.: Evaluation of classification trees for fast segmentation of white matter fiber tracts. In: Proc. MICCAI wksp. Computational Diffusion MRI (2008)

Design and Construction of a Realistic DWI Phantom for Filtering Performance Assessment

Antonio Tristán-Vega and Santiago Aja-Fernández

Laboratory of Image Processing, University of Valladolid. Spain
atriveg@lpi.tel.uva.es, sanaja@tel.uva.es

Abstract. A methodology to build a realistic phantom for the assessment of filtering performance in Diffusion Weighted Images (DWI) is presented. From a real DWI data–set, a regularization process is carried out taking into account the diffusion model. This process drives to a model which accurately preserves the structural characteristics of actual DWI volumes, being in addition regular enough to be considered as a *noise–free* data–set and therefore to be used as a ground–truth. We compare our phantom with a kind of simplified phantoms commonly used in the literature (those based on homogeneous cross sections), concluding that the latter may introduce important biases in common quality measures used in the filtering performance assessment, and even drive to erroneous conclusions in the comparison of different filtering techniques.

1 Introduction

Diffusion Tensor Imaging (DTI) allows the study of the fiber tracts in the white matter of the brain *in vivo*. DTI are obtained from Diffusion Weighted Images (DWI); each of them is acquired by applying a sensitizing gradient in a known direction, which produces an attenuation in the T2 (baseline) image following an exponential law [1]. By collecting six or more independent gradient directions, the underlying fiber structure at each voxel may be inferred in terms of the six free components of the Diffusion Tensor (DT) [2]. As opposed to conventional MRI, DWI show a poor Signal to Noise Ratio (SNR), since the signal power is lower due to the DWI attenuation. This is especially the case with modern High Angular Resolution Diffusion Imaging (HARDI) techniques, where very strong gradients (and therefore very strong attenuation) have to be applied in order to improve the angular contrast [3]. Moreover, the Rician nature of the noise in DWI [4] prevents the use of conventional Gaussian-based filtering techniques.

The adverse effect of Rician noise in DWI has been previously reported in [5,6], where it has been shown that it produces artifacts which hinder the recovery of the original signal once the DT has been estimated. To palliate the effect of noise, a number of techniques may be used, including regularization after DT estimation [7], regularized estimation of the DT [8] and DWI restoration before DT estimation. In this last category, a number of techniques have been proposed: the Conventional Approach [9] (CA), based on the properties of the second order moment of Rician data; Maximum Likelihood (ML) estimation [10]; anisotropic diffusion [11]; wavelets [12]; total variation [13]; Unbiased Non Local Means (UNLM)

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 951–958, 2009.

[14,15]; Linear Minimum Mean Squared Error (LMMSE) filtering [16,17]; multi–channel, Rician–corrected Wiener filtering [18], and others.

The quantitative assessment of filtering performance has been carried out in different ways. They include visual assessment [14,15,16,17] and indirect measures based on the properties of the DTI volumes recovered from filtered DWI [16,18]; in this case, the computation of related parameters such as the smoothness in the fiber tracts estimated from DTI [18] make it difficult to evaluate the filtering performance, since the final result depends on a number of factors others than the filtering itself. On the other hand, direct evaluation on the filtered DWI is a difficult task; in [14] 12 acquisitions of the same patient are available, so a leave–one–out strategy is used to filter one of the volumes each time and compare the result to a *noise–free* image obtained from the remaining 11 volumes. When multiple observations are not available, synthetic data has to be used. As opposed to conventional MRI, for which very realistic phantoms exist and have been intensively used [19], there is not a DWI synthetic data–set widely accepted as a standard, so over-simplified models are used like in [15,16,17], based on large homogeneous regions simulating coarse crossing fiber tracts. Commonly used configurations are crossing sections in two [15,16,18] and three [17] dimensions, although other approaches like the "Earth" and "logarithm" in [18] or the logmarithmic spiral in [16] are possible. Nevertheless, these simple configurations are not representative of the complex architecture of the white matter of the brain, which in general presents multiple fiber crossings, bending and sharp changes in curvature/orientation in the space of a few voxels. Thus, we propose a novel methodology to design a highly realistic DWI phantom to be used as a gold standard for filtering performance assessment. It is built from a real DWI data–set, so it is able to account for the high complexity of the human brain, but at the same time we use a number of regularization techniques which allow us to obtain a nearly *noise–free* signal which may be used as a ground truth. Additionally, we empirically show that commonly used phantoms may bias the results on performance assessment, driving to erroneous conclusions. We illustrate this by the comparison of two popular DWI filtering schemes (UNLM and LMMSE) based on a conventional phantom and the one here introduced.

2 Materials

The phantom has been built from a real DWI volume of a volunteer. We use a SENSE EPI data–set scanned in a 3 Tesla General Electrics Echospeed system. (Sequence: Maximum gradient amplitudes: 40 mT/M. Rectangular Field of view of 220 x 165 mm. Slice thickness 1.7 mm. Receiver bandwidth: 6kHz. TE 70 ms; TR 80 ms (effective TR 2500 ms)). It comprises 8 non–weighted baseline images and 51 gradient directions, and is $256 \times 256 \times 81$ voxels in size, with a resolution of $0.94 \times 0.94 \times 1.7$ mm^3. The b value is 586 s/mm^2; although this value is quite small for practical applications (typical values are greater than 1000 s/mm^2), it has the advantage that its SNR is also greater.

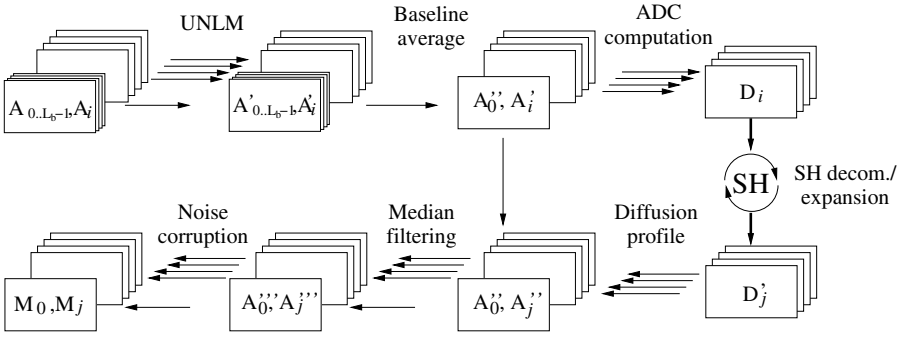


Fig. 1. Overview of the construction of the synthetic phantom. The input is the original (real) DWI volume, and the output the synthetic volume.

3 Method

3.1 Overview

Fig. 1 summarizes the steps on the design of the phantom, which we detail later on. We filter the 8 baseline images ($A_0 \dots A_{L_b-1}$, with $L_b = 8$) and the 51 gradient directions ($A_{L_b} \dots A_{L_b+L_g-1}$, with $L_g = 51$) to reduce the amount of noise ($A'_0 \dots A'_{L_b+L_g-1}$). The 8 baselines are then averaged to obtain one single, nearly *noise–free* baseline (A''_0). All gradient directions are divided by the baseline to obtain the attenuation signal, and from it the Apparent Diffusion Coefficient (ADC) by taking its logarithm ($D_1 \dots D_{L_g}$). We use a Spherical Harmonics representation to regularize the ADC. This smoothed ADC ($D'_1 \dots D'_{L'_g}$), together with the *noise–free* baseline A''_0 is used to produce synthetic gradient directions ($A''_1 \dots A''_{L'_g}$). The resulting DWI volume is median filtered ($A'''_0 \dots A'''_{L'_g}$) to remove outliers and obtain the *noise–free* ground truth. The input to the filters will be the noise–corrupted phantom ($M_0 \dots M_{L'_g}$).

3.2 Filtering and Baseline Average

Although the low b value of the DWI volume produces a relatively high SNR, a previous filtering is needed to drastically reduce noise. We use the popular UNLM as described in [15], with a search radius of 5 voxels, a comparison radius of 2 voxels, and $h = \sigma$. Assuming that the residual noise in the baselines after filtering is of zero mean, we average $A'_0 \dots A'_{L_b-1}$ to obtain A''_0, which is assumed to be *noise–free*. The gradient images have lower SNR, so they need further regularization taking into account the diffusion model, as explained next.

3.3 Computation of the Apparent Diffusion Coefficient

Under the assumption of Gaussian diffusion, the (true) diffusion signal E_i is related to the (true) baseline E_0 by the Stejskal–Tanner equation [1]:

$$E_i = E_0 \exp\left(-b \cdot g_i^T \mathcal{D} g_i\right), \; i = 1 \dots L_g \tag{1}$$

with g_i the i-th direction and $\mathcal{D}$ the diffusion tensor. However, this equation cannot model the diffusion process in the event of complex configurations such as fiber crossing or bending [3]. In this case, the diffusion may be represented in terms of a positive function D defined on the unit sphere, the ADC:

$$E_i = E_0 \exp\left(-b \cdot D\left(g_i\right)\right) \Rightarrow D\left(g_i\right) \simeq -\log\left(A_i'/A_0''\right)/b,\ i = 1 \ldots L_g \tag{2}$$

3.4 Analysis in the Basis of Spherical Harmonics

Since the ADC is defined on the unit sphere, it may be represented in the basis of Spherical Harmonics. We use the regularized least squares approach in [3] with $\lambda = 0.006$ and order 6 (28 basis functions), which in our experiments showed the best trade–off between representation capability (i.e. it does not hinder the ability to resolve crossings or alter the angular contrast) and regularization. This way, the *noise–free* (due to least squares regularization) ADC is:

$$D'(g) = \textstyle\sum_{l=0}^{27} C_l Y_l(g) \tag{3}$$

where Y_l are the basis functions and C_l are the coefficients of the expansion obtained from $D(g_i)$ in eq. (2) (see [3] for details). Note that the basis functions $Y(g)$ are continuous–valued. Therefore, once we have chosen an arbitrary set of gradient directions g_j', $j = 1 \ldots L_g'$, the *noise free* diffusion signal is, see eq. (2):

$$A_j'' = A_0'' \exp\left(-b' \textstyle\sum_{l=0}^{27} C_l Y_l(g_j')\right),\ j = 1 \ldots L_g' \tag{4}$$

The set of gradient directions g_j' and the magnitude of the sensitizing gradients b' are not necessarily the same as the original ones, g_i and b. In fact, we use a more realistic value for practical applications, $b' = 1200$ s/mm^2. For g_j, we generate 15 gradient directions as antipodal pairs.

3.5 Median Filtering

As a final processing, all gradient images and the baseline are filtered slice–by–slice with a median filter of size 3×3. This step is required to remove isolated outliers due to a poor fit of the spherical harmonics. Although very few voxels are affected by this artifact, this is useful to achieve a visually adequate result.

3.6 Noise Corruption

To produce the inputs to the filter, we corrupt the image with Rician noise [4]:

$$M_j = \sqrt{\left(A_j''' + n_c\right)^2 + n_s^2},\ j = 0 \ldots L_g' \tag{5}$$

where n_c and n_s are independent Gaussian random processes with zero mean and power σ^2. The noise in MRI is usually not white, but it is correlated inside each slice due to the spatial interpolation produced by the zero padding of the

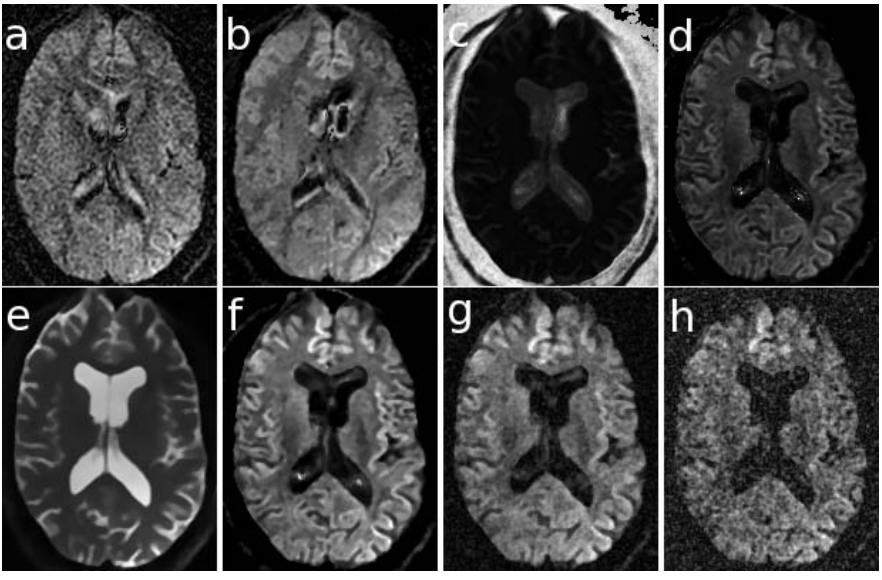


Fig. 2. Illustration of the whole process of phantom construction; from the original gradients A_i (**a**), UNLM is used to drastically reduce noise (A'_i in **b**). Then the ADC (D_i) is computed and regularized (D'_j) via spherical harmonics decomposition, with a total of 28 coefficients C_l (**c** shows the first one). Arbitrary gradient directions A''_j are generated with a higher b value (**d**). Finally, a median filter is used to obtain the *noise–free* phantom A'''_j, comprising the averaged baseline A'''_0 (**e**) and each of the 15 gradient images A'''_j (**f**). Correlated noise is added to produce the noisy phantoms M_i, with a maximum SNR of 54 dB (**g**) and a minimum of 48 dB (**h**) (Original: 51.7dB).

spectrum of the image in the k–space. To account for this characteristic, we generate white noise processes $\tilde{n}_{c,s}$ and low–pass filter them inside each slice with a 2–D Gaussian kernel $G(p_1, p_2)$ in the spatial (image) domain. It has an isotropic variance η^2:

$$n_{c,s}(p_1, p_2) = \left(\sum_{q_1,q_2} G^2(q_1, q_2) \right)^{-1} \sum_{q_1,q_2} \tilde{n}_{c,s}(q_1, q_2) G(p_1 - q_1, p_2 - q_2) \qquad (6)$$

where dividing by the energy of G is necessary to keep the noise power constant.

4 Results and Discussion

We show in Fig. 2 an illustration of the whole process described in Section 3. First, note that although UNLM is able to achieve very clean images (compare **a** and **b**), the filtered image is not an adequate phantom: there is some residual noise, and the contrast of the image is very poor. This justifies the need for the regularization of the ADC (**c** and **d**) and the increase in b. As a result, the

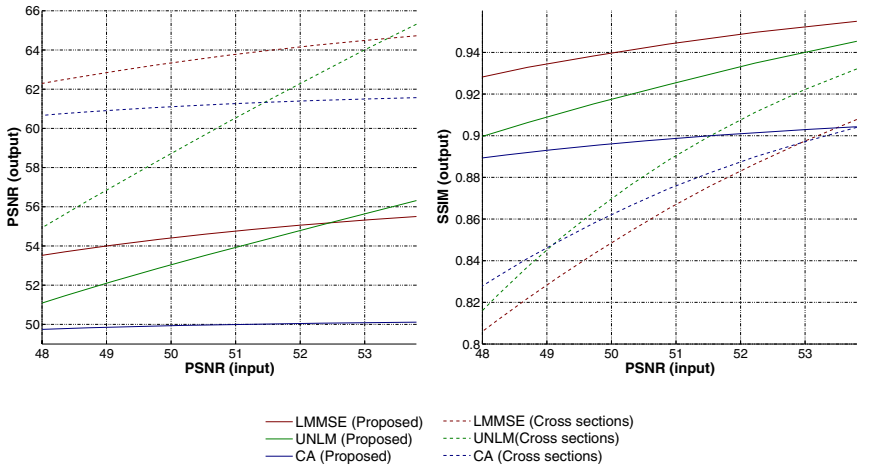


Fig. 3. Filtering performance (PSNR and SSIM) for Linear Minimum Mean Squared Error (LMMSE), Unbiased Non–Local Means (UNLM) and Conventional Approach (CA) filters, for a simple cross sections phantom and the proposed phantom. Results are shown for 15 gradient directions.

synthetic gradient images (**d**) are cleaner and have a realistic contrast. On the contrary, the baseline image obtained from UNLM filtering and averaging (**e**) has an excellent quality, since the original baselines are far less noisy than the original gradient images. Although the synthetic gradients are very clean, some outliers may be appreciated in **d**, so median filtering is used to obtain the final phantom **f**. Nevertheless, note that the majority of outliers are located in the background and the CSF (where no fiber bundles are present), which should be removed for the assessment of filtering performance. The final result (**e** and **f**) is a very clean image, but yet it has well defined borders, being an appropriate phantom. In **g** and **h** we show noise corrupted images, in a high SNR[1] (**g**) and a low SNR case (**h**). Note that practically any realistic situation will lie between these two extreme cases. For the correlation filter, we use $\eta^2 = 0.8^2$, estimated from the original image (**a**), so the noisy spots accurately resemble those in **a**.

To show the importance of an adequate phantom, we have intended an illustrative example. We are going to numerically compare different filtering algorithms for DWI, using the phantom here presented and a more typical phantom based on cross sections; the phantom is a clone of the 3D cross described in [17], but we have used 15 gradient directions and scaled the magnitude of the baseline and the gradient images so that they all show mean values equal to the mean values of our own phantom. Regarding the algorithms, we have chosen two recent approaches: the popular UNLM implemented as described in [15], with the

[1] We use the Peak Signal to Noise Ratio (PSNR) (maximum SNR in the baseline image) to give numeric results in all cases.

parameters proposed in Section 3.2, and the LMMSE filter described in [17] with the code issued by the authors and the parameters suggested in [17]. Additionally, we test the Conventional Approach (CA), see [9], implemented as an isotropic Gaussian low–pass filter with $\eta = 1.5$ applied to the squared signal, followed by bias removal and computation of the squared root. As similarity measures, we use the PSNR and the Structural Similarity (SSIM) index [20] between the filtered volume and the ground truth (see Fig. 2–**f**). Results may be found in Fig. 3. CA, although being a very simple approach, shows better results than UNLM for PSNR (for low input SNR) and than LMMSE for SSIM (practically for all input SNR) when using the cross sections phantom. When we use our realistic phantom, CA obviously shows worse results than the other filters, since it is not designed to preserve the borders. In the cross sections phantom there are practically no borders, so this is not an issue. If we center our attention in the SSIM index, it is worth notice that the two phantoms give completely opposite results for UNLM and LMMSE for all input PSNR; the realistic phantom shows that LMMSE shows a better behavior than UNLM for this particular index, but using the over–simplified phantom would lead to the erroneous conclusion that UNLM yields better SSIM indices. Although this behavior cannot be generalized from this single experiment, it is useful to show the bias that an unrealistic phantom may induce.

5 Conclusions

A methodology to build a realistic DWI phantom to assess the accuracy of filtering algorithms for DWI has been described. Contrary to other validation methods proposed in the literature, see [14], our methodology does not require multiple observations of the same volume; the phantom may be built from one single DWI data–set. Moreover, the same real–data volume may be used to produce phantoms with very different configurations (different b values and numbers of gradient directions) as described in Section 3.4, so it may be used to test filtering algorithms in a wide variety of situations. On the other hand, we have shown an example of how an inadequate phantom may lead to erroneous conclusions in the assessment of filtering performance, which justifies the importance of the work presented.

Acknowledgments. Authors would like to acknowledge grant number TEC2007–67073/TCM from the Comisión Interministerial de Ciencia y Tecnología (Spain).

References

1. Stejskal, E.-O., Tanner, J.-E.: Spin diffusion measurements: Spin echoes in the presence of a time-dependent field gradient. J. of Chem. Phys. 42, 288–292 (1965)
2. Basser, P.J., Pierpaoli, C.: Microstructural and physiological features of tissues elucidated by Quantitative–Diffusion–Tensor MRI. J. of Magn. Res. 111(3), 209–219 (1996)

3. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical Q-Ball imaging. Magn. Res. in Med. 58, 497–510 (2007)
4. Gudbjartsson, H., Patz, S.: The Rician distribution of noisy MRI data. Magn. Res. in Med. 34, 910–914 (1995)
5. Basser, P.J., Pajevic, S.: Statistical artifacts in diffusion tensor MRI (DT–MRI) caused by background noise. Magn. Res. in Med. 44, 41–50 (2000)
6. Jones, D., Basser, P.J.: Squashing peanuts and smashing pumpkins: how noise distorts diffusion–weighted MR data. Magn. Res. in. Med. 52, 979–993 (2004)
7. Pennec, X., Fillard, P., Ayache, N.: A Riemannian framework for tensor computing. Intl. J. of Com. Vision 66, 41–66 (2006)
8. Fillard, P., Pennec, X., Arsigny, V., Ayache, N.: Clinical DT–MRI estimation smoothing, and fiber tracking with log–Euclidean metrics. IEEE Trans. on Med. Imag. 26, 1472–1482 (2007)
9. McGibney, G., Smith, M.: Unbiased signal to noise ratio measure for magentic resonance images. Med. Phys. 20(4), 1077–1078 (1993)
10. Sijbers, J., den Dekker, A.-J.: Maximum Likelihood estimation of signal amplitude and noise variance from MR data. Magn. Res. Imag. 51, 586–594 (2004)
11. Parker, G.J., Schnabel, J.A., Symms, M.R., Werring, D.J., Barker, G.J.: Nonlinear smoothing for reduction of systematic and random errors in diffusion tensor imaging. J. of Magn. Res. Imag. 11, 702–710 (2000)
12. Pižurica, A., Philips, W., Lemahieu, I., Acheroy, M.: A versatile Wavelet domain noise filtration technique for medical imaging. IEEE Trans. on Med. Imag. 22(3), 323–331 (2003)
13. McGraw, T., Vemuri, B.C., Chen, Y., Mareci, T., Rao, M.: DT–MRI denoising and neural fiber tracking. Med. Imag. Anal. 8, 95–111 (2004)
14. Wiest-Daesslé, N., Prima, S., Coupé, P., Morrissey, S.P., Barillot, C.: Non-local means variants for denoising of diffusion-weighted and diffusion tensor MRI. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 344–351. Springer, Heidelberg (2007)
15. Wiest-Daesslé, N., Prima, S., Coupé, P., Morrissey, S.P., Barillot, C.: Rician noise removal by Non–Local Means filtering for low signal–to–noise ratio MRI: Applications to DT–MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 171–179. Springer, Heidelberg (2008)
16. Aja-Fernández, S., Niethammer, M., Kubicki, M., Shenton, M.E., Westin, C.-F.: Restoration of DWI data using a Rician LMMSE estimator. IEEE Trans. on Med. Imag. 27(10), 1389–1403 (2008)
17. Tristán-Vega, A., Aja-Fernández, S.: Joint LMMSE estimation of DWI data for DTI processing. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 27–34. Springer, Heidelberg (2008)
18. Martín-Fernández, M., Muñoz–Moreno, E., Cammoun, L., Thiran, J.–P., Westin, C.–F., Alberola-López, C.: Sequential anisotropic multichannel Wiener filtering with Rician bias correction applied to 3D regularization of DWI data. Med. Imag. Anal. 13, 19–35 (2009)
19. Collins, D.L., Zijdenbos, A.P., Kollokian, V., Sled, J.G., Kabani, N.J., Holmes, C.J., Evans, A.C.: Design and construction of a realistic digital brain phantom. IEEE Trans. on Med. Imag. 17(3), 463–468 (1998)
20. Wang, Z., Bovik, A.-C., Sheikh, H.-R., Simoncelli, E.-P.: Image quality assessment: form error visibility to structural similarity. IEEE Trans. on Im. Process. 13(4), 600–612 (2004)

Statistical Detection of Longitudinal Changes between Apparent Diffusion Coefficient Images: Application to Multiple Sclerosis*

Hervé Boisgontier[1,2], Vincent Noblet[1], Félix Renard[1,2], Fabrice Heitz[1], Lucien Rumbach[2,3], and Jean-Paul Armspach[2]

[1] Université de Strasbourg, CNRS, UMR 7005, LSIIT, France
[2] Université de Strasbourg, CNRS, UMR 7191, LINC-IPB, France
[3] Service de Neurologie, CHU Minjoz, France
h.boisgontier@unistra.fr

Abstract. The automatic analysis of longitudinal changes between Diffusion Tensor Imaging (DTI) acquisitions is a promising tool for monitoring disease evolution. However, few works address this issue and existing methods are generally limited to the detection of changes between scalar images characterizing diffusion properties, such as Fractional Anisotropy or Mean Diffusivity, while richer information can be exploited from the whole set of Apparent Diffusion Coefficient (ADC) images that can be derived from a DTI acquisition. In this paper, we present a general framework for detecting changes between two sets of ADC images and we investigate the performance of four statistical tests. Results are presented on both simulated and real data in the context of the follow-up of multiple sclerosis lesion evolution.

1 Introduction

The automatic analysis of longitudinal changes between diffusion tensor imaging (DTI) acquisitions is a promising tool for monitoring neurodegenerative disease and particularly for the follow-up of patients suffering from Multiple Sclerosis (MS) [1]. Indeed, studies have already highlighted diffusion property alterations induced by MS by comparing Fractional Anisotropy (FA) and/or Mean Diffusivity (MD) values in some manually selected regions of interest (ROI) [2,3,4]. However, few works address the issue of automatically detecting changes between DTI acquisitions. Recently, a statistical framework based on nonparametric permutation testing has been proposed to automatically detect significant changes between scalar images characterizing diffusion properties [5]. An alternative parametric statistical test, which relies on the Generalized Likelihood Ratio Test, have also been proposed to detect changes in DTI-derived scalar images [6].

* We would like to thank the ARSEP (Association pour la Recherche sur la Sclérose En Plaques) and the Région Alsace for their support.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 959–966, 2009.

The main limitation of these previous approaches is that they can only detect changes between scalar images such as FA or MD, while richer information can be obtained from the whole set of Apparent Diffusion Coefficient (ADC) images that can be derived from a DTI acquisition. Indeed, a DTI acquisition is composed of a T_2-weighted image (without diffusion weighting) and a set of N diffusion weighted (DW) images, from which can be derived N ADC images characterizing the diffusion properties in N directions. These images have the potential to finely characterize brain microstructure organization. Here, we present a general framework for detecting changes between two sets of ADC images. This framework is composed of preprocessings that aim at registering the two DTI acquisition in a common coordinate system and at correcting them from undesirable geometrical distortions, and of a statistical detection step. In this paper, we investigate the performance of four statistical tests [7]. First, we consider the independent t-test that tests if the two sets of ADC represent samples drawn according to Gaussian distributions with different mean values. The three other tests under investigation assumed that the two sets of ADC are dependent samples, that means that for a given direction, the ADC of the two acquisitions are correlated. This assumption makes sense since the ADC profile is representative of brain tissue organization. Under the hypothesis that no change occurs, ADC profiles of two acquisitions of the same subject mapped in a common coordinate system should correlate well for each voxel. Thus, we consider a parametric test (dependent t-test) and two nonparametric tests (sign test and Wilcoxon signed-ranks test) dedicated to dependent samples. Since the two acquisitions may have been obtained according to different gradient sampling schemes and since ADC profiles may require to be reoriented when registering DTI acquisition [8], the ADC values of the two exams may not be sampled according to same directions. To make the use of dependent tests valid, we resample the ADC values of one exam according to the same gradient directions as the second one. To this end, we resort to an appropriate spherical interpolation method that extends standard linear interpolation to data sampled on a spherical grid [9]. Results are presented on both simulated and real data in the context of the follow-up of multiple sclerosis lesion evolution.

2 Method

2.1 General Framework for Change Detection

The goal of change detection is the following: given two longitudinal DTI acquisitions $\mathcal{A}_1$ and $\mathcal{A}_2$, decide for each voxel whether a significant change has occurred. The proposed detection scheme is composed of the following steps: (i) eddy-current distortion correction of $\mathcal{A}_1$ and $\mathcal{A}_2$, (ii) registration of $\mathcal{A}_1$ onto $\mathcal{A}_2$, (iii) resampling of ADC images of $\mathcal{A}_1$ according to gradient directions of $\mathcal{A}_2$ (required only when considering dependent statistical test), and (iv) statistical detection test. Eddy-current distortion correction is achieved by affinely registering each slice of the DW-images onto the corresponding slice of the T_2-weighted image using mutual information.

The two ADC images datasets are then affinely registered, the transformation being estimated by considering the mutual information between the FA image of the two acquisitions. Such transformation may not always be sufficient to compensate all the undesired global differences that may be present between two acquisitions. These remaining changes, that are not of interest, may be the consequence of additional geometrical distortion or of brain atrophy. Thus, a deformable registration method is required. We use the topology preserving B-spline based registration method presented in [10], the estimation of the transformation being also done by considering the FA images. In order not to remove the changes of interest, which are supposed to be focal changes, we only consider a transformation with 10 125 parameters.

For resampling the ADC values of $\mathcal{A}_1$ according to gradient directions of $\mathcal{A}_2$, we resort to an appropriate spherical interpolation scheme that extends the standard linear interpolation scheme to data sampled on a spherical grid [9]. Indeed, ADC values can be considered as data sampled on a spherical grid, the location on the sphere being given by the corresponding gradient direction. Thus, an interpolated value is computed as a weighted sum of its three nearest neighbor values, the weights being computed by considering the areas of spherical triangles calculated thanks to the L'Hullier's theorem [9].

Finally, statistical change detection is achieved.

2.2 Statistical Detection of Changes between ADC Images

Let S_0 be the image without diffusion weighting, S_i the i^{th} DW image characterizing diffusion in gradient direction $\left[g_{i;x}, g_{i;y}, g_{i;z}\right]^t$, b the diffusion weighting factor and $N \geq 6$ the number of diffusion directions. Let $Y = [1/b\ \log(S_0/S_1), \cdots, 1/b\ \log(S_0/S_N)]^t$ be the corresponding ADC. Let $Y_1(\boldsymbol{v})$ and $Y_2(\boldsymbol{v})$ be the ADC values at a given voxel $\boldsymbol{v}$ of acquisitions $\mathcal{A}_1$ and $\mathcal{A}_2$ respectively. $Y_1(\boldsymbol{v})$ and $Y_2(\boldsymbol{v})$ are assumed to be of dimension N_1 and N_2 respectively. After the resampling step, required for dependent tests, $Y_1(\boldsymbol{v})$ and $Y_2(\boldsymbol{v})$ are assumed to be of the same dimension N.

In our experiments, N is equal to 33. This is a somewhat limited number of samples to allow using the normal approximation in most of the tests presented in the sequel. This is why, for statistical testing at a given voxel, we also consider the ADC values of the 26 neighbors. Thus, this is both a mean of increasing the statistical power of the tests and a way to account for spatial information.

Among the broad panel of existing statistical tests, we choose to consider the independent t-test, the dependent t-test, the sign test and the Wilcoxon signed-ranks test. These statistical tests are briefly summarized in the sequel. Notice that all the considered tests are able to evaluate whether two sets of ADC are different in average (two-tailed hypothesis: $\mu_1 \neq \mu_2$), of whether one set of ADC is in average higher or lower than the other one (one-tailed hypothesis: *e.g.* $\mu_1 > \mu_2$). In our experiments, we are interested in all kinds of changes (two-tailed hypothesis). Thus, the statistical detection map are built by considering the absolute value of the statistical test, so that change areas can be obtained just by an upper thresholding.

Independent Student's *t*-test: The independent Student's t-test [7] evaluates the following hypothesis: do the two sets of ADC, assumed to be independent measures with unequal variance, represent samples drawn according to Gaussian distributions with different mean values ? The t statistic to test has the following expression:

$$t = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}, \tag{1}$$

where μ_1, μ_2 and σ_1^2, σ_2^2 are the sample mean and the sample variance of the ADC values of acquisitions $\mathcal{A}_1$ and $\mathcal{A}_2$ respectively. The distribution of t can then be approximated by a Student's t distribution with the degrees of freedom calculated using:

$$dof = \frac{\left(\sigma_1^2/N_1 + \sigma_2^2/N_2\right)^2}{\left(\sigma_1^2/N_1\right)^2/(N_1-1) + \left(\sigma_2^2/N_2\right)^2/(N_2-1)}. \tag{2}$$

Dependent Student's *t*-test: The dependent Student's t-test [7] evaluates the following hypothesis: do the sets of differences between corresponding ADC values in the two acquisitions represent samples drawn according to a Gaussian distribution with zero mean value ? The t statistic to test can be calculated as follows:

$$t = \frac{\mu_d}{\sqrt{\frac{\sigma_d^2}{N}}}, \tag{3}$$

where μ_d and σ_d^2 designate the sample mean and the sample variance of the sets of differences between paired ADC values. In that context, t is assumed to follow a Student's distribution with $N - 1$ degrees of freedom.

Sign Test: The sign test [7] is a simple nonparametric method to compare two dependent samples. It is based on the distribution of the signs of the differences observed between paired samples, whatever the amount of these differences. The sign test evaluates the following hypothesis: are there equal numbers of positive and negative differences in the samples ? *i.e.* $\mathcal{H}_0$: $P(+) = P(-) = \frac{1}{2}$. We consider the normal approximation involving a correction for continuity of the sign test, which can be written as follows:

$$Z = \frac{|N^+ - N| - 0.5}{\frac{\sqrt{N}}{2}}, \tag{4}$$

where N^+ is the number of positive differences between paired ADC values. Under these assumptions, Z follows a normal distribution.

Wilcoxon Matched-Pairs Signed-Ranks Test: The Wilcoxon matched-pairs signed-ranks test [7] is an extension of the sign test. Contrary to the sign test, it takes into account the amount of the differences, which leads to a more

powerful test. The absolute differences between paired ADC values are arranged in ascending order. The sign is then re-attached to each rank. Under null hypothesis, the sum of the ranks with a negative sign must be roughly equal to the sum of the ranks with a positive sign. We consider the normal approximation of the Wilcoxon signed-ranks test, which can be written as follows:

$$Z = \frac{T - N(N+1)/4}{\sqrt{N(N+1)(2N+1)/24}}, \tag{5}$$

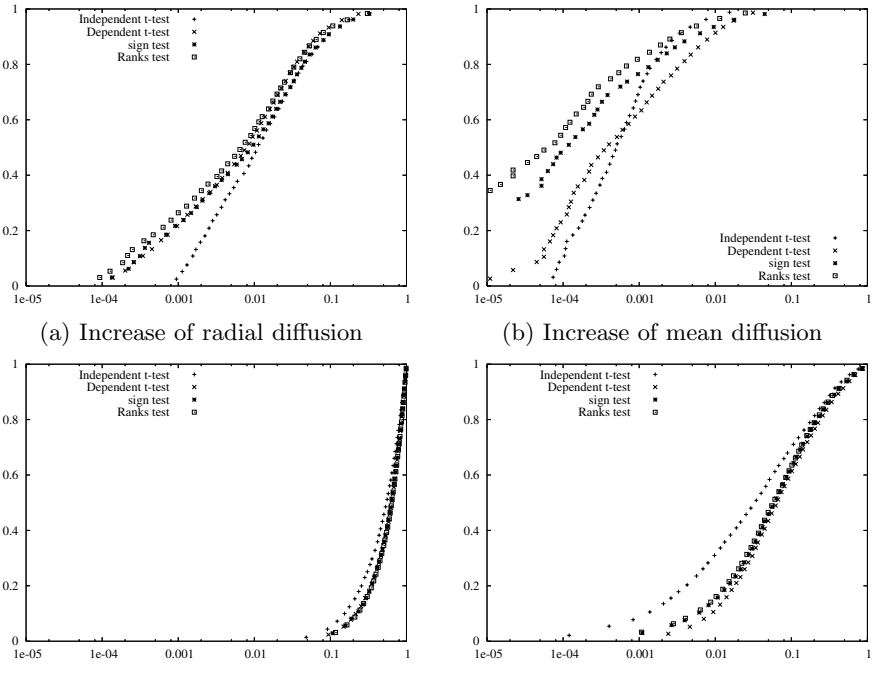
where T is the sum of the ranks with the less frequent sign. Under these assumptions, Z follows a normal distribution.

3 Experimental Results

3.1 Experiments on Simulated Data

To assess the proposed approach, we first resort to synthetic simulations. Four kinds of changes have been simulated: (i) increase of radial diffusion, (ii) increase of mean diffusion, (iii) change of diffusion orientation and (iv) decrease of longitudinal diffusion. To this end, we consider the DTI acquisition (33 directions) of a healthy subject (the baseline) and a second DTI dataset obtained the next day from the same subject. By this way, we account for undesired changes that may appear between two acquisitions (change in patient positioning, drift of acquisition parameters between scans, acquisition artifacts, noise) and which may possibly increase the rate of false detections. Lesion evolutions are then simulated in several ROI of the second dataset. For a realistic simulation, the ROI have been defined by segmenting lesions in an image from a patient suffering from MS and by transporting them in the coordinate system of the healthy subject using a non-rigid registration method [10]. Since simulating decrease of longitudinal diffusion is relevant only in areas of anisotropic diffusion, we consider for this very simulation only regions with FA higher than 0.7. Modifications of longitudinal, radial and mean diffusions are achieved by first estimating tensors and corresponding eigenvalues, and then by modifying appropriately these eigenvalues and propagating these changes to DW images [6]. Modification of diffusion orientation is done by rotating the gradient directions $\left[g_{i;x}, g_{i;y}, g_{i;z}\right]^t$ in the ROI. Areas between regions of change and areas with no change are modified in order to obtain smooth transitions.

To compare the performance of the four statistical tests, ROC (Receiver Operating Characteristic) curves have been plotted by considering several thresholds. The results presented in Figure 1 highlight that the different tests achieve detection with a performance that vary with the kind of simulated changes. For simulated changes in the radial diffusion (Fig. 1a), the three paired tests lead roughly to same results and outperform the independent t-test. For detecting changes of the mean diffusivity (Fig. 1b), the Wilcoxon matched-pairs signed-ranks test is the best performer. The results of the sign test are slightly worse but better compared to the dependent t-test. The independent t-test, which leads



(a) Increase of radial diffusion

(b) Increase of mean diffusion

(c) Change of diffusion orientation

(d) Decrease of longitudinal diffusion

Fig. 1. ROC curves comparing detection results on different kinds of simulated changes with the sign test, the ranks test, the dependent t-test and the independent t-test

to the worst results for high detection thresholds, yields results similar to the signed-ranks test for lower thresholds. Concerning simulated changes of diffusion orientation, all the tests reveal to be very inefficient since results are not better than if they were due to chance. This can easily be explained for the independent t-test since modification of diffusion orientation does not change the mean ADC value, thus leaving the result of this test unchanged. For the other three tests, it turns out that changes in diffusion orientation induce in average a similar amount of positive and negative differences, thus having less impact in the evaluation of these tests. Finally, detecting decrease in longitudinal diffusion (Fig. 1d) is best achieved with the independent t-test. The three paired methods all perform roughly with the same performance. Such kind of modification only affects a limited number of ADC values (only those whose direction is almost colinear with the principal diffusion direction), thus being quite difficult to detect because of a lack of statistical power. In summary, the three paired tests show similar results, with a little better performance for the ranks test and a slightly poorer performance for the dependent t-test. The independent t-test shows a quite different behavior, generally leading to worse results except for the detection of decreases of longitudinal diffusion.

3.2 Follow-Up on MS Patients

The different methods have been tested for an MS patient that underwent two MR exams at one year distance. An expert manually labelled evolving lesions on the T_2-weighted images between these two exams. We compare the segmentation from the expert with the detections obtained with the four tests (Fig. 2 and 3). The ROC curve presented in Fig. 3 shows that the three paired tests lead to almost similar results, while the independent t-test is significantly better. Notice that this ROC curve should be analyzed carefully since the segmentation from the expert cannot really be considered as a ground truth (changes may occur in diffusion imaging without being visible in T_2-weighted images.).

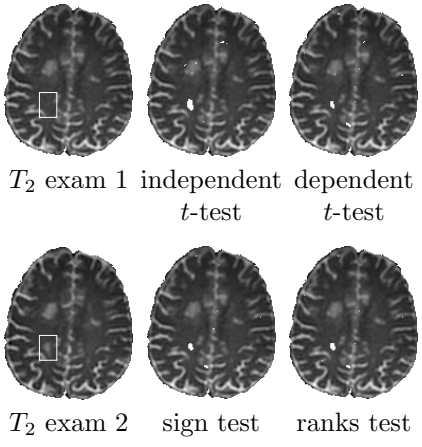


Fig. 2. Results obtained with the four tests on a real case

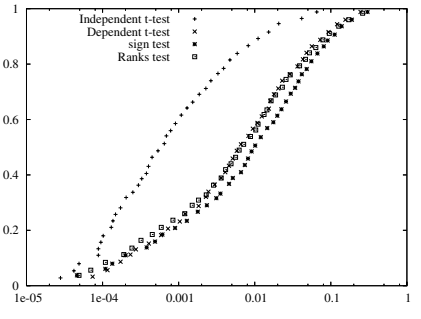


Fig. 3. ROC curves comparing detection results with the four tests on a real case

When analyzing qualitatively the detections (Fig. 2: detections have been obtained by selecting a threshold, according to the ROC curves presented in Fig. 3, that allows 0.1 % of false detections), we can observe that most lesion evolutions visible in the T_2-weighted image have also been detected by the four statistical tests (Fig. 2), except one (not shown here), corresponding to a decrease of longitudinal diffusion combined with an increase of radial diffusion and with an unchanged mean diffusivity.

4 Conclusion

We have proposed a framework to compare two sets of ADC and investigated the performance of four statistical tests. To the best knowledge of the authors, this is the first work that investigate to detect longitudinal changes between DTI acquisitions by using statistical hypothesis testings on ADC images. Results on simulated and real data highlighted different behaviors of these tests. Among the paired tests, the ranks test seems to always lead to slightly better results than the two other tests. However, the comparison between independent and

paired tests did not enable to point out a best performer, the ranks test being more effective on simulated data for detecting increase of radial diffusion and mean diffusivity while the independent t-test showed a better performance for detecting decrease of longitudinal diffusion and also for detecting lesion evolution on the real case. This conclusion seems at first sight quite surprising since one can expect that a paired test is more sensitive for detecting changes. This is in fact the case, the main limitation being that such tests are also more sensitive for detecting unwilling changes induced, for instance, by registration errors. The independent t-test appears to be more robust, thus leading to better results on the real scenario.

An important issue has been ignored in this work, which is how to automatically choose the threshold to get the significant detections. Although all the statistical tests under investigation yield p-values maps, defining a threshold while controlling the false detection rate is not an easy task because of the multiple comparisons problem and also because of the spatial correlation of the data. Further work will be done to investigate this issue.

References

1. Ge, Y.: Multiple Sclerosis: The Role of MR Imaging. American Journal of Neuroradiology 27(6), 1165–1176 (2006)
2. Werring, D.J., Brassat, D., Droogan, A.G., Clark, C.A., Symms, M.R., Barker, G.J., MacManus, D.G., Thompson, A.J., Miller, D.H.: The pathogenesis of lesions and normal-appearing white matter changes in multiple sclerosis: a serial diffusion MRI study. Brain 123(8), 1667–1676 (2000)
3. Cassol, E., Ranjeva, J.P., Ibarrola, D., Mékies, C., Manelfe, C., Clanet, M., Berry, I.: Diffusion tensor imaging in multiple sclerosis: a tool for monitoring changes in normal-appearing white matter. Multiple Sclerosis 10(2), 188–196 (2004)
4. Garaci, F.G., Colangelo, V., Ludovici, A., Gaudiello, F., Marziali, S., Centonze, D., Boffa, L., Simonetti, G., Floris, R.: A diffusion longitudinal MR imaging study in normal-appearing white matter in untreated relapsing-remitting multiple sclerosis. American Journal of Neuroradiology 28, 475–478 (2007)
5. Chung, S., Pelletier, D., Sdika, M., Lu, Y., Berman, J.I., Henry, R.G.: Whole brain voxel-wise analysis of single-subject serial DTI by permutation testing. NeuroImage 39(4), 1693–1705 (2007)
6. Boisgontier, H., Noblet, V., Heitz, F., Rumbach, L., Armspach, J.P.: An automatic method for change detection in serial DTI-derived scalar images. In: MICCAI workshop on MIAMS, New-York, USA (September 2008)
7. Sheskin, D.: Handbook of parametric and nonparametric statistical procedures. Chapman and Hall/CRC, Boca Raton (2003)
8. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. IEEE Transactions on Medical Imaging 20(11), 1131–1139 (2001)
9. Francesca, C.M.: Interpolation on spherical geodesic grids: A comparative study. Journal of Computational and Applied Mathematics 210(1-2), 99–105 (2007)
10. Noblet, V., Heinrich, C., Heitz, F., Armspach, J.P.: 3-D deformable image registration: a topology preservation scheme based on hierarchical deformation models and interval analysis optimization. IEEE Transactions on Image Processing 14(5), 553–566 (2005)

Tensor-Based Analysis of Genetic Influences on Brain Integrity Using DTI in 100 Twins

Agatha D. Lee[1], Natasha Leporé[1], Caroline Brun[1], Yi-Yu Chou[1], Marina Barysheva[1], Ming-Chang Chiang[1], Sarah K. Madsen[1], Greig I. de Zubicaray[2], Katie L. McMahon[2], Margaret J. Wright[3], Arthur W. Toga[1], and Paul M. Thompson[1]

[1] Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Los Angeles, CA
[2] Functional MRI Laboratory, Centre for Magnetic Resonance, University of Queensland, Brisbane, Australia
[3] Queensland Institute of Medical Research, Brisbane, Australia

Abstract. Information from the full diffusion tensor (DT) was used to compute voxel-wise genetic contributions to brain fiber microstructure. First, we designed a new multivariate intraclass correlation formula in the log-Euclidean framework [1]. We then analyzed used the full multivariate structure of the tensor in a multivariate version of a voxel-wise maximum-likelihood structural equation model (SEM) that computes the variance contributions in the DTs from genetic (A), common environmental (C) and unique environmental (E) factors. Our algorithm was tested on DT images from 25 identical and 25 fraternal twin pairs. After linear and fluid registration to a mean template, we computed the intraclass correlation and Falconer's heritability statistic for several scalar DT-derived measures and for the full multivariate tensors. Covariance matrices were found from the DTs, and inputted into SEM. Analyzing the full DT enhanced the detection of A and C effects. This approach should empower imaging genetics studies that use DTI.

1 Introduction

Imaging genetics is a rapidly growing field that combines mathematical methods from genetics and imaging to identify factors that contribute to variations in brain structure and function. A more mechanistic understanding of brain structure could be achieved if the contributing genes could be identified. Each gene's contribution is expected to be minor, so researchers have first looked for genetically influenced measures in images in the hope of finding specific features or brain regions in which gene effects are the strongest. Twin studies, in particular, examine statistical differences between pairs of identical and fraternal twins; as these twins share all or half their genes on average, structural equation models may be fitted to understand what proportion of the variance in image-derived measures is attributed to genetic differences.

Genetic factors influence various aspects of brain anatomy including cortical thickness [16], regional gray and white matter volumes [4], and univariate measures of fiber integrity derived from DTI [4,10]. Even so, few studies have analyzed genetic influences on signals that are inherently multidimensional, such as diffusion tensors.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 967–974, 2009.

By reducing the full DT to scalar measures of diffusion anisotropy, potentially relevant information is thrown away. DTI measures the multidirectional profile of water diffusion in brain tissue, revealing information on brain architecture and composition. In most DTI studies, a diffusion tensor (DT) is fitted to model at each voxel; its eigenvalues represent the magnitude of diffusion along three orthogonal principal directions - fiber directions may be inferred from the principal eigenvector. The local fractional anisotropy (FA), computed from the eigenvalues, is widely accepted as an index of fiber integrity and is correlated with intellectual performance [5]. An alternative anisotropy measure, the geodesic anisotropy (GA) [3], measures the geodesic distance between tensors on the symmetric positive-definite tensor manifold. In [9], we found that a multivariate statistical analysis of the full diffusion tensor outperformed derived scalar signals in detecting group differences in the blind; others also suggest that effect sizes in group DTI studies can be greater when retaining the full tensor information [17].

Here we present a new approach that compute differences and correlations between DTs using the "Log-Euclidean" framework. Our new intraclass correlation is derived from a Mahalanobis-like distance in the SPD(3) manifold of symmetric positive definite matrices. DTs that correspond to a physically possible diffusion process must lie in the SPD(3) Lie group, and do not form a vector subspace of the vector space of matrices with the usual matrix addition and scalar multiplication. To account for this, we performed all statistical computations in the Log-Euclidean framework [2], which allows standard Euclidean computations on the DT manifold. We used this manifold distance to compute 3D whole-brain maps of heritability using the classical Falconer method [6], and we compared heritability maps from the full DT to maps based on standard univariate measures (including FA, GA and tGA).

We also generalized the A/C/E model used in quantitative genetics, to assess genetic influences on multidimensional signals such as DTI. This method enables us to assess genetic influences in brain architecture by computing variances within and between members of twin pairs, using distances computed on the diffusion tensor manifold. We compared multivariate and scalar DTI-derived measures to determine the most heritable measures. Honing in on heritable measures is typically the first step in identifying specific genes that affect brain structure and function [7].

2 New Multivariate Statistical Formulae for Twin Study

2.1 Intra-class Correlations on the Tensor Manifold

The intraclass correlation [14] between pairs of observations is defined, for univariate quantities such as FA, as:

$$r_{uni} = \frac{MS_{between} - MS_{within}}{MS_{between} + MS_{within}}, \tag{1}$$

where $MS_{between}$ and MS_{within} are mean-square estimates of the between-pair and within-pair variance, respectively.

For tensor-valued data such as the full 3 x 3 DT, equation (1) cannot be used. Instead we use a distance in the tensor manifold measuring the deviation of subject's diffusion tensor T_i from the mean of the entire sample, weighted by the variances:

$$r_{multi} = \frac{\frac{1}{Np}\sum_{i=1}^{Np}((\log T_1^i - \log \bar{T})^T I(\log T_2^i - \log \bar{T}))}{\sigma_x \cdot \sigma_y}, \tag{2}$$

where $\sigma_x = \sqrt{\frac{1}{Np}\sum_{i=1}^{Np}((\log T_1^i - \log \bar{T})^T I(\log T_1^i - \log \bar{T}))}$ and $\sigma_y = \sqrt{\frac{1}{Np}\sum_{i=1}^{Np}((\log T_2^i - \log \bar{T})^T I(\log T_2^i - \log \bar{T}))}$, $\bar{T}$ is the mean of a set of vectors T^i, i=1,...,m, $\boldsymbol{T_1}^i$ and $\boldsymbol{T_2}^i$ represent the log-transformed tensors for each member of the *i*-th twin pair (for simpler notation, we consider the 6 unique components of the log-transformed tensor as belonging to a 6-component Euclidean vector). Here, we reshaped 3x3 diffusion tensor matrix to 6x1 matrix as diffusion tensor is symmetric giving only 6-variate data. We define $\bar{T} = \exp(\frac{1}{M}\sum_{i=1} m \log T^i)$ where M is the total number of subjects, N_p is number of twin pairs and I is the 6x6 identity tensor. I could be omitted, but is included to show that it could be generalized to an inverse covariance matrix expressing the empirical correlation between the 6 unique DT components at each voxel (we reserve this generalization for future work).

2.2 Structural Equation Modeling (A/C/E model) of Variance in DTI Volume

To understand the relative contribution of additive genetic (A) versus shared (C) and unique (E) environmental effects on the variance in a DTI-derived measure, for example the FA or GA, we compute the measure at each voxel in a set of MZ pairs and DZ pairs and measure the covariance between twin 1 and twin 2 at each voxel. These empirically estimated covariance matrices can be computed for any observed variable (Z), and a structural equation model (SEM) can be fitted to the covariances to infer the proportion of the variance attributable to each of A, C and E. Measurement errors or inter-subject registration errors will both be classified as part of the E component of variance. Z for one twin pair may be modeled as:

$$Z = aA + cC + eE\,. \tag{3}$$

where A/C/E are latent variables and *a, c, e* are the weights of each parameter to be determined.

We used a maximum-likelihood estimate (MLE) [12] to estimate the proportion of the voxel-based intersubject variance that is attributable to each of the 3 free model parameters. The 3 variance components combine to create the total observed inter-individual variance, so that $a^2+c^2+e^2=1$. The weights $\theta = (a, c, e)$ are estimated by comparing the covariance matrix implied by the model, $\Sigma(\theta)$, and the sample covariance matrix of the observed variables, S, using maximum-likelihood fitting:

$$F_{ML,\theta} = \log|\Sigma(\theta)| + trace(\Sigma^{-1}(\theta)\mathbf{S}) - \log|\mathbf{S}| - p, \tag{4}$$

where $p = 2$ is the number of observed variables. Under the null hypothesis that Z is multivariate normal (i.e., each of A, C and E are normally distributed), the MLE model follows a χ^2 distribution with $p(p+1) - t$ degrees of freedom, where t is number of model parameters (3 in our case). We used the Broyden-Fletcher-Goldfarb-Shanno method [13] to obtain the minimum $F_{ML,\theta}$. This algorithm is used for both univariate and multivariate data, the only difference being in the computation of the components of the covariance matrix S.

2.3 Illustration of the Covariance Computations on the Manifold

For the univariate data such as FA, the covariance matrix is defined as:

$$Cov(t_1, t_2) = \frac{1}{Np}\sum_{i=1}^{Np}(t_1^i - \bar{t}_1)(t_2^i - \bar{t}_2) \cdot \tag{5}$$

where $\bar{t}$ is the mean of a set t^i i=1,...,m, $\boldsymbol{t_1}$ and $\boldsymbol{t_2}$ represent each subject of the twin pair, $\bar{t} = \exp(\frac{1}{M}\sum_{i=1} m \log t^i)$.

In the case of the 3-component vector whose elements are the eigenvalues, or the full DT, there are either 3 or 6 parameters per voxel, all containing potentially useful information for genetic analysis. The covariance matrices can no longer be computed using the previous general formula. Here we propose a new multivariate generalization of Eq. 5 in the Log-Euclidean formalism (Fig 1). In the Log-Euclidean formalism, distances are computed in the tangent space at the origin of the manifold of positive-definite, symmetric matrices. This plane is reached via a matrix logarithm. Hence, for multivariate measures such as 3x3 full DT, which has been reshaped to 6x1 matrix for the computational purpose, the covariance equation becomes:

$$Cov(T_1, T_2) = \frac{1}{Np}\sum_{i=1}^{Np}((\log T_1^i - \log \bar{T})(\log T_2^i - \log \bar{T})^T) \tag{6}$$

This distance is illustrated in **Fig 1**, where each twin's tensors are represented as points on the curved manifold. This distance may be thought of as a bilinear form that takes two tensors as arguments and returns their discrepancy. As noted earlier, the

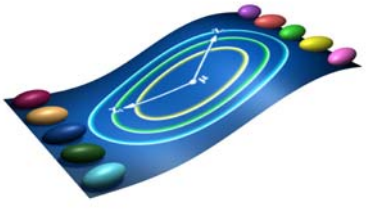


Fig. 1. Here we define a distance between two tensors in the tensor manifold, SPD(3), as $(x_1 - \mu)^T(x_2 - \mu)$, where μ is the mean tensor of all the twins in log-Euclidean space

distances in the manifold could be generalized to take into account the naturally occurring (and perhaps genetically mediated) correlations between the tensor component *i* in twin 1 and tensor component *j* in twin 2. For simplicity, we use the standard metric on Log-Euclidean space, not a statistical metric.

3 Implementation

To test our analysis methods, we acquired 3D structural brain MRI and DTI scans from 100 subjects: 25 identical (MZ) twin pairs (25.1±1.5SD yrs old) and 25 fraternal (DZ) twin pairs (23.1±2.1 yrs) on a 4 Tesla Bruker Medspec MRI scanner with an optimized diffusion tensor sequence. Imaging parameters were: 21 axial slices (5 mm thick), FOV = 23 cm, TR/TE 6090/91.7 ms, 0.5 mm gap, with a 128x100 acquisition matrix. 30 gradient images were collected: three scans with no diffusion sensitization (i.e., T2-weighted images) and 27 diffusion-weighted images for which gradient directions were evenly distributed on the hemisphere [7]. The reconstruction matrix was 128x128, yielding a1.8x1.8 mm^2 in-plane resolution.

3.1 Image Preprocessing and Registration

3D structural MR images were automatically skull-stripped using the Brain Surface Extraction software (BSE) [15] followed by manual editing, and registered via 9-parameter affine transformation to a high-resolution single-subject brain template image, the Colin27 template, using the FLIRT software [8]. 3D structural images were registered to a Mean Deformation Template (MDT) using a 3D fluid registration [11]. Jacobian matrices were obtained from the resulting deformation fields.

DTs were computed from the diffusion-weighted images and smoothed using Log-Euclidean tensor denoising to eliminate singular, negative definite, or rank-deficient tensors, using MedINRIA (http://www.sop.inria.fr/asclepios/software/MedINRIA). To eliminate extracerebral tissues, non-brain tissues were manually deleted from one of the diagonal component images (D_{xx}), yielding a binary brain extraction mask (cerebellum included). Masked images were registered by 12-parameter transformation to the corresponding 3D structural images in the standard template space using *FLIRT* [8].

Transformation parameters from affine and nonlinear registrations were used to rotationally reorient the tensors at each voxel [1] to ensure that the multidimensional tensor orientations remained consistent with the anatomy after image transformation [1,18]. We used two separate algorithms to compute the tensor rotations: the finite strain (FS) and the preservation of principal direction (PPD) algorithms ([1,18])

3.2 Statistical Analysis for Twins

We computed FA, GA and tGA values and the matrix logarithms of the full diffusion tensors for each subject. GA is the manifold equivalent of the FA in the Log-Euclidean framework [2,10]:

$$GA(S) = \sqrt{Trace(\log S - < \log S > I)^2}\,. \tag{7}$$

with $< \log S >= \frac{Trace(\log S)}{3}$.

We used the hyperbolic tangent of GA, *tGA* as in [3], to create maps with a comparable range to the *FA*, i.e., [0,1]. Two sets of voxel-wise covariance matrices for the MZ pairs and DZ pairs were computed for all the univariate and multivariate measures detailed above. For each statistic, we estimated the intraclass correlation, heritability (i.e., variance proportion due to genetic differences among individuals), and computed the best-fitting A/C/E model. For intraclass correlations, we computed permutation-based *p* values to assess their significance.

4 Results and Discussion

To examine the intra-pair variance for each type of twin, Fig 2 shows intraclass correlation (*r*) maps between MZ pairs and DZ pairs for FA (Fig 2a and e), GA (Fig 2b and f), tGA (Fig 2c, g), and the full multivariate DT (Fig 2d, h).

Supporting the validity of these measures, the ventricles, corpus callosum and some anterior temporal regions show higher resemblance among MZs than among DZ twins for all anisotropy measures. Maps based on DT-derived scalars (FA, GA, tGA) are relatively noisy, but maps from the full multivariate tensor show higher SNR than the univariate analyses. Maps of Falconer's heritability estimate for FA, GA, tGA, and the full DT matrix (Fig 2i-l) show that the fiber characteristics of the corpus callosum and some anterior temporal regions are heritable (as confirmed statistically in the p(A/E) maps of Fig 3), consistent with prior studies that only examined FA [5]. Heritability maps confirm the correlation patterns seen in the MZ twins, but adjust for the correlations in DZ twins.

Contributions of factor A and C are shown in Fig 3a-h for the scalar measures (FA, GA, tGA) and the multivariate full DT. Probability maps based on voxel-wise chi-squared statistics confirm the A/C/E model's goodness of fit for all measures. We stress that in A/C/E and other structural equation models, a probability of less than 0.05 indicates that the model is *not* a good fit, so values of p>0.05 are the values of interest (contrary to the usual case in brain maps).

Importantly, maps based on the full diffusion tensor have higher p-values, which means the models are a better fit, and they also appear more spatially coherent. The A and C factors are higher too, suggesting that there is less unmodeled residual variance (which is lumped into the E term).

The p-values from SEM are corrected for the multiple comparisons using FDR. pFDR values for each of the DT-derived measures are pFDR=0.038 for the FA, pFDR=0.0234 for the GA, pFDR=0.038 for the tGA, pFDR=0.0408 for the full DT, indicating that the A/C/E model is a good fit after multiple comparisons correction.

In the *r*-value maps (Fig 2), maps generated using multivariate data (Fig 3d and h) have higher SNR than the univariate data (Fig 3a-c and e-g). Our genetic maps based on the full DT afford better fitting of genetic and environmental models than scalar indices that discard relevant information. These highly heritable phenotypes should facilitate the quest for single-gene polymorphisms that influence the fiber architecture

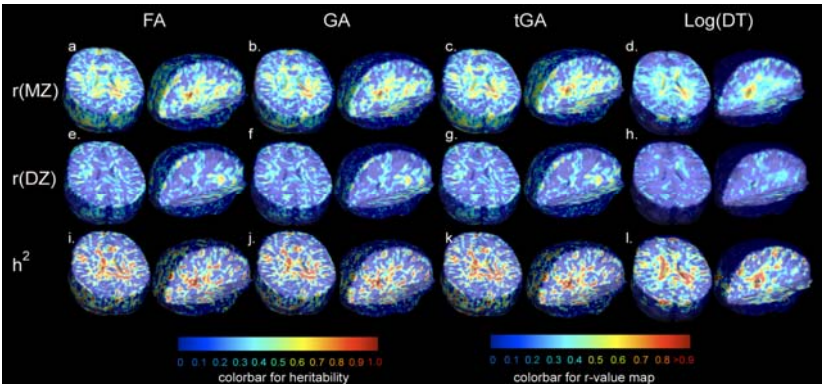


Fig. 2. (**a-h**) show intraclass correlation maps for FA, GA**,** tGA and the full DT

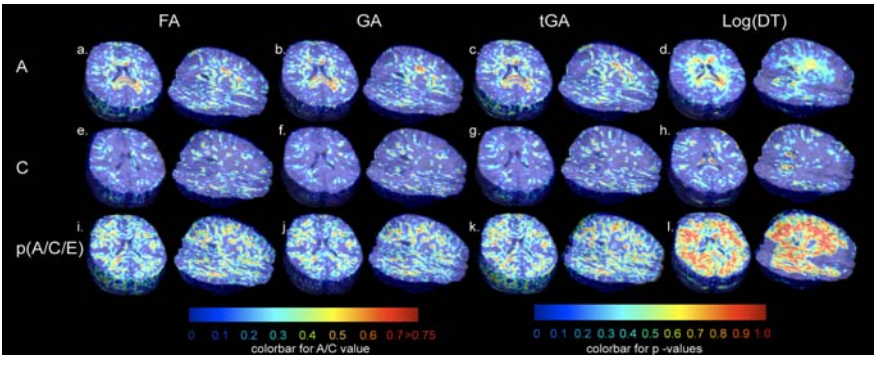


Fig. 3. Maps **(a-d)** show genetic (A) and **(e-h)** shared environmental (C) proportions of variance for various DTI-derived measures: FA, GA**,** tGA and the full DT. The goodness of fit of the A/C/E genetic model is confirmed at voxels where p exceeds 0.05 in panels **(i-l).** In general, the full DT shows best effect sizes for fitting a genetic model; the anisotropy indices show a moderately good fit.

of the living brain, as they hone in on measures and specific brain regions where genetic influences are the most powerfully detected.

The resulting maps show the expected phenotypic patterns of genetic and environmental influences. The ventricles, corpus callosum and anterior temporal regions that develop earlier in life showed strong genetic influences even after multiple comparisons correction. Multivariate analysis also offered improved SNR.

References

1. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. IEEE-TMI 20, 1131–1139 (2001)
2. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Fast and Simple Calculus on Tensors in the Log-Euclidean Framework. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 115–122. Springer, Heidelberg (2005)

3. Batchelor, P., Moakher, M., Atkinson, D., Calamante, F., Connelly, A.: A rigorous framework for diffusion tensor calculus. Magn. Reson. Med. 53, 221–225 (2005)
4. Brun, C.A., Leporé, N., Pennec, X., Chou, Y.-Y., Lee, A.D., Barysheva, M., de Zubicaray, G., Meredith, M., McMahon, K.L., Wright, M.J., Toga, A.W., Thompson, P.M.: A tensor-based morphometry study of genetic influences on brain structure using a new fluid registration method. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 914–921. Springer, Heidelberg (2008)
5. Chiang, M.C., Barysheva, M., Lee, A.D., Madsen, S., Klunder, A.D., Toga, A.W., McMahon, K.L., Zubicaray, G.I., Meredith, M., Wright, M.J., Srivastava, A., Balov, N., Thompson, P.M.: Mapping genetic influences on brain fiber architecture with HARDI. In: ISBI, pp. 871–874 (2008)
6. Falconer, D.S.: Introduction to Quantitative Genetics, 2nd edn. Longman, Harlow (1981)
7. Glahn, D.C., Thompson, P.M., Blangero, J.: Neuroimaging endophenotypes: Strategies for finding genes influencing brain structure and function. Hum. Br. Mapp. 28(6), 488–501 (2007)
8. Jenkinson, M., Smith, S.: A global optimisation method for robust affine registration of brain images. Med. Image. Anal. 5, 143–156 (2001)
9. Lee, A.D., Lepore, N., Lepore, F., Alary, F., Voss, P., Brun, C., Chou, Y.-Y., Barysheva, M., Toga, A.W., Thompson, P.M.: Brain Differences Visualized in the Blind Using Tensor Manifold Statistics and Diffusion Tensor Imaging. In: FBIT 2007, pp. 470–476 (2007)
10. Lee, A.D., Lepore, N., Barysheva, M., Brun, C., Chou, Y.-Y., Madsen, S., McMahon, K.L., Zubicaray, G.I., Wright, M.J., Toga, A.W., Thompson, P.M.: Comparison of fractional and geodesic anisotropy in monozygotic and dizygotic twins from diffusion tensor imaging. In: IEEE ISBI, pp. 943–946 (2008)
11. Lepore, N., Brun, C.A., Chiang, M.-C., Chou, Y.-Y., Dutton, R.A., Hayashi, K.M., Lopez, O.L., Aizenstein, H.J., Toga, A.W., Becker, J.T., Thompson, P.M.: Multivariate statistics of the jacobian matrices in tensor based morphometry and their application to HIV/AIDS. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 191–198. Springer, Heidelberg (2006)
12. Neale, M.C., Boker, S.M., Xie, G., Maes, H.H.: Mx: Statistical modeling (1999)
13. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical recipes in C, Cambridge (1994)
14. Scout, P.E., Fleiss, J.L.: Intraclass correlations: Uses in assessing rater reliability. Psych. Bull. 2, 420–428 (1979)
15. Shattuck, D.W., Leahy, R.M.: BrainSuite: an automated cortical surface identification tool. Medical Image Analysis 8(202), 129–141 (2001)
16. Thompson, P.M., Cannon, T.D., Narr, K.L., et al.: Genetic influences on brain structure. Nat. Neuro. 1253–1258 (2001)
17. Verma, R., Davatzikos, C.: Manifold based analysis of diffusion tensor images using isomaps. Biomedical Imaging: Nano to Macro. 790–793 (2006)
18. Zhang, H., Yushkevich, P.A., Alexander, D.C., Gee, J.C.: Deformable registration of diffusion tensor MR images with explicit orientation optimization. Medical Image Analysis 10, 764–785 (2006)

Robust Extrapolation Scheme for Fast Estimation of 3D Ising Field Partition Functions: Application to Within-Subject fMRI Data Analysis*

Laurent Risser[1,2,3], Thomas Vincent[1,2], Philippe Ciuciu[1,2], and Jérôme Idier[3]

[1] NeuroSpin/CEA, 91191 Gif-sur-Yvette, France
[2] IFR 49, Institut d'Imagerie Neurofonctionnelle, Paris, France
[3] IRCCyN/CNRS, 1 rue de la Noë 44300 Nantes, France

Abstract. In this paper, we present a fast numerical scheme to estimate Partition Functions (PF) of 3D Ising fields. Our strategy is applied to the context of the joint detection-estimation of brain activity from functional Magnetic Resonance Imaging (fMRI) data, where the goal is to automatically recover activated regions and estimate region-dependent hemodynamic filters. For any region, a specific binary Markov random field may embody spatial correlation over the hidden states of the voxels by modeling whether they are activated or not. To make this spatial regularization fully adaptive, our approach is first based upon a classical path-sampling method to approximate a small subset of *reference* PFs corresponding to prespecified regions. Then, the proposed extrapolation method allows us to approximate the PFs associated with the Ising fields defined over the remaining brain regions. In comparison with preexisting approaches, our method is robust to topological inhomogeneities in the definition of the *reference* regions. As a result, it strongly alleviates the computational burden and makes spatially adaptive regularization of whole brain fMRI datasets feasible.

1 Introduction

In fMRI, one usually resorts to spatial filtering to enhance the signal-to-noise ratio at the expense of a loss of spatial resolution. A more challenging approach works on the unsmoothed data by introducing some prior knowledge on the sought spatial structures through for instance local interaction models such as Markov Random Fields (MRFs). Discrete MRFs, which have been used in segmentation and clustering, typically involve a set of hyper-parameters: the smaller this number the less complex the patterns modelled by the corresponding MRF. For instance, a single temperature level controls the amount of spatial correlation in *symmetric* Ising fields. In the considered fMRI application [1], such Ising fields are hidden since the activation detection process is modelled a priori through a

* We are thankful to Région Ile de France for funding.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 975–983, 2009.

two-class Spatial Mixture Model (SMM). Moreover, their definition varies within a brain parcellation that segregates the 4D data into Γ functionally homogeneous non-regular parcels ($\Gamma \approx 500$) in order to take the regional fluctuations of the Hemodynamic Filter (HF) into account. The Joint Detection-Estimation (JDE) analysis thus consists of a parcel-based analysis where Γ independent HFs as well as stimulus-dependent SMMs are identified in order to detect the regions involved in the paradigm and their underlying dynamics (see Section 2). Therefore, several hundreds of temperature levels have to be estimated making a hand-tuning procedure unrealistic. Moreover, since the optimal setting of such a parameter may be different when considering different parcels, all parameters cannot be fixed to the same value. The purpose of this paper is then to provide an unsupervised scheme for adapting the amount of spatial regularization in several hundreds of Ising fields with different topological configurations.

For a single Ising field, estimating the temperature hyper-parameter requires a precise estimation of its Partition Function (PF)[1] since the exact PF computation is intractable (see Section 3). This estimation can be achieved using importance sampling techniques [2,3]. Recently, alternative strategies have been introduced in the literature [4,5]. In short, they resort to *continuous* MRFs and *logistic* transform to build up an approximation that maps continuous weights to the binary latent variables in a mixture model. While efficient, these approaches require the setting of an additional scale parameter that controls the accuracy of the approximation and whose optimal value may depend on the application. Moreover, these approximation techniques have been validated only for MRFs defined over regular grids. Hence, we propose in Section 3 a fast and robust extrapolation technique to *multiple* PF estimation. In comparison to preexisting approaches, our method is robust against grid inhomogeneities and efficient irrespective of the parcel configurations. Validation on real fMRI data is perfomed in Section 4. In Section 5, we discuss the pros and cons of our approach, which has applications in image processing beyond the fMRI context.

2 Joint Detection-Estimation of Brain Activity in fMRI

The JDE framework proposed [1,6] relies on a prior parcellation of the brain into $\boldsymbol{\mathcal{P}} = (\mathcal{P}_\gamma)_{\gamma=1:\Gamma}$ functionally homogeneous and connected parcels [7]. Every parcel $\mathcal{P}_\gamma$ comprising voxels $(V_j)_{j=1:J}$ is characterized by a region-based model of the BOLD signal, which consists in estimating a single HF shape $\boldsymbol{h}_\gamma$ whatever the number of experimental conditions involved in the paradigm.

In a given parcel $\mathcal{P}_\gamma$, voxel-dependent and stimulus-induced fluctuations of the BOLD signal are encoded by spatially varying magnitudes $\mathbb{a} = (a_j^m)_{j=1:J}^{m=1:M}$, where m stands for the stimulus type index. The fMRI time course measured in voxel $V_j \in \mathcal{P}_\gamma$ then reads: $\boldsymbol{y}_j = \sum_{m=1}^{M} a_j^m \boldsymbol{x}^m \star \boldsymbol{h}_\gamma + \boldsymbol{b}_j$, where $\boldsymbol{x}^m$ stands for the m^{th} binary stimuli vector and $\boldsymbol{b}_j$ stands for the serially correlated noise component [1, 6]. Within the Bayesian framework, prior probability density

[1] The normalizing constant that makes the MRF measure of unit mass over its domain.

functions (pdfs) are introduced on every sought object *i.e.*, $(\mathbb{a}, \boldsymbol{h}_\gamma)$ [1]. Spatial Gaussian mixture models are expressed on $\mathbb{a}$ through the introduction of hidden variables $\mathbb{q} = (\boldsymbol{q}_j^m)_{j=1:J}^{m=1:M}$ that encode whether voxel V_j is activating in response to stimulus m ($q_j^m = 1$) or not ($q_j^m = 0$). Hence, stimulus-dependent hidden symmetric Ising fields are introduced on these states:

$$\Pr(\boldsymbol{q}^m \,|\, \beta_m) = Z(\beta_m)^{-1} \exp\left(\beta_m U(\boldsymbol{q}^m)\right), \tag{1}$$
$$\text{with} \quad Z(\beta_m) = \sum_{\boldsymbol{q}^m \in \{0,1\}^n} \exp\left(\beta_m\, U(\boldsymbol{q}^m)\right)$$

and $U(\boldsymbol{q}^m) = \sum_{i\sim j} I(q_i^m = q_j^m)$ is the global "negative energy" where $I(A) = 1$ whenever A is true and 0 otherwise. The parameter $\beta_m \geq 0$ controls the amount of spatial correlation between the components of $\boldsymbol{q}^m$ according to the grid $\mathcal{G}$, while $Z(\beta_m)$ defines the partition function. The global prior SMM then reads:

$$p(\mathbb{a} \,|\, \Theta_{\mathbb{a}}) = \prod_m p(\boldsymbol{a}_m \,|\, \boldsymbol{\theta}_m) = \prod_m \Big\{ \sum_{\boldsymbol{q}^m} \Big[\prod_j f(a_j^m \,|\, q_j^m, \boldsymbol{\theta}_m) \Big] \Pr(\boldsymbol{q}^m \,|\, \beta_m) \Big\}$$

where $f(a_j^m \,|\, q_j^m = i) \sim \mathcal{N}(\mu_{i,m}, v_{i,m})$. Parameters $\mu_{i,m}$ and $v_{i,m}$ define the prior mean and variance of class $i = 0, 1$, respectively for the stimulus type m. Let the set $\boldsymbol{\theta}_m$ gather mixture parameters $\{\mu_{0,m}, \mu_{1,m}, v_{0,m}, v_{1,m}, \beta_m\}$. Samples of the full posterior pdf $p(\boldsymbol{h}_\gamma, \mathbb{a}, \mathbb{q}, \boldsymbol{\Theta} \,|\, \mathbb{y})$ are simulated using a Gibbs sampler algorithm and posterior mean estimates are then computed from these samples. Here, we specifically deal with the sampling of parameter β_m, which is achieved using a *symmetric* random walk Metropolis-Hastings step: at iteration k, a candidate $\beta_m^{(k+1/2)} \sim \mathcal{N}(\beta_m^{(k)}, \sigma_\epsilon^2)$ is generated. It is accepted (*i.e.*, $\beta_m^{(k+1)} = \beta_m^{(k+1/2)}$) with probability: $\alpha(\beta_m^{(k)} \to \beta_m^{(k+1/2)}) = \min(1, A_{k,k+1/2})$, where the acceptation ratio $A_{k,k+1/2}$ follows from Eq. (1):

$$A_{k,k+1/2} = \frac{p(\beta_m^{(k+1/2)} | \boldsymbol{q}_m^{(k)})}{p(\beta_m^{(k)} | \boldsymbol{q}_m^{(k)})} = \frac{p(\boldsymbol{q}_m^{(k)} | \beta_m^{(k+1/2)}) p(\beta_m^{(k+1/2)})}{p(\boldsymbol{q}_m^{(k)} | \beta_m^{(k)}) p(\beta_m^{(k)})}$$
$$= \frac{Z(\beta_m^{(k)})}{Z(\beta_m^{(k+1/2)})} \exp\left((\beta_m^{(k+1/2)} - \beta_m^{(k)}) U(\boldsymbol{q}_m^{(k)}) \right),$$

using Bayes' rule and considering a uniform prior for β_m. The β_m sampling step then requires to estimate ratios of $Z(.)$ or log-PF differences for all $\mathcal{P}_\gamma$ parcels prior to exploring the full posterior pdf. This motivates the need for developing fast approximation techniques of these constants.

3 Ising Field PF Estimation

3.1 Single PF Estimation

Exact evaluation of $Z(\cdot)$ in a reasonable amount of time is impossible except on tiny grids. Robust and fast estimation of $Z(\beta)$ is thus a key issue for numerous 3D imaging problems involving Ising models and more generally discrete MRFs.

Path-sampling is an extension of importance sampling for estimating ratios of normalizing constants, by considering series of easy-to-sample unormalized intermediate densities. Such a strategy was proven efficient to tabulate the PF for the Ising case; see [8] for details. Time dedicated to a single PF estimation using this method is acceptable but becomes penalizing when numerous PFs need to be estimated as required when dealing with several hundreds of grids of variable size and shape. Since this typical situation occurs in our fMRI application, a fast compromise consists in resorting to path-sampling to get log-scale estimates $(\log \widehat{Z}_{\mathcal{G}_p}(\beta))_{p=1:P}$ for a small subset of *reference* grids $(\mathcal{G}_p)_{p=1:P}$ and then in using extrapolation formulas to obtain $\log \widetilde{Z}_{\mathcal{T}}(\beta)$ for the large remaining set of brain parcels to be analyzed, referenced here by a *test* grid $\mathcal{T}$.

3.2 Multiple PF Estimation

In [8], the authors address linear regression on $(\log \widehat{Z}_{\mathcal{G}_p}(\beta))_{p=1:P}$ as a function of the number of cliques in the grids $(\mathcal{G}_p)_{p=1:P}$. Estimates $\log \widetilde{Z}_{\mathcal{T}}(\beta)$ are then linearly computed from the regression coefficients and the number of cliques in $\mathcal{T}$. A bilinear extension of this technique, which also accounts for the number of sites in the grid, can be thought of to estimate log-PFs in *small and non-regular grids*[2] such as those appearing in our fMRI application. However, it can be shown that the accuracy of regression-based techniques strongly depends on the homogeneity and the number of *reference* grids: the more inhomogeneous the reference set, the larger the approximation error. These reasons motivate the development of more reliable approaches regarding the reference PFs.

3.3 Proposed PF Extrapolation Technique

The appropriate reference grid is selected by means of a similarity measure and the accuracy of the approximation is controlled by an error criterion. Our algorithm proceeds in two steps: 1) Akin to [8], reference PFs $\widehat{Z}_{\mathcal{G}_p}(\beta_k)$ are estimated using path-sampling. The topological configurations of the reference grids $(\mathcal{G}_p)_{p=1:P}$ can now be inhomogeneous to cover a maximum of grid configurations that may occur in further PF estimations. 2) For any test grid $\mathcal{T}$, $\log Z_{\mathcal{T}}$ is approximated using a *single* reference log-PF estimate out of $(\log \widehat{Z}_{\mathcal{T}_p}(\beta))_{p=1:P}$.

Let $r_{\mathcal{T}} = \sigma_{n,\mathcal{T}}/\mu_{n,\mathcal{T}}$ be a measure of grid homogeneity where $\mu_{n,\mathcal{T}}$ and $\sigma_{n,\mathcal{T}}$ respectively provide the average number of neighbors per site over $\mathcal{T}$ and the corresponding standard deviation. The smaller $r_{\mathcal{T}}$ the more regular $\mathcal{T}$. Using this criterion, our similarity measure is defined by $\mathcal{L}_{\mathcal{T}}(\mathcal{G}_p) = \|r_{\mathcal{T}} - r_{\mathcal{G}_p}\|^2$, which helps us choose the closest reference grid $\mathcal{G}_{\text{ref}}$ to $\mathcal{T}$ in combination with the approximation error criterion $\mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ given by:

$$\mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p) = \|\log Z_{\mathcal{T}}(\beta) - \log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_p)\|^2 / \|\log Z_{\mathcal{T}}(\beta)\|^2 \qquad (2)$$

$$\text{where} \quad \log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_p) = \big(c_{\mathcal{T}}(\log \widehat{Z}_{\mathcal{G}_p}(\beta) - \log 2)/c_{\mathcal{G}_p} + \log 2\big). \qquad (3)$$

[2] Here, non-regular grids make reference to regular lattices combined with non-straight boundaries.

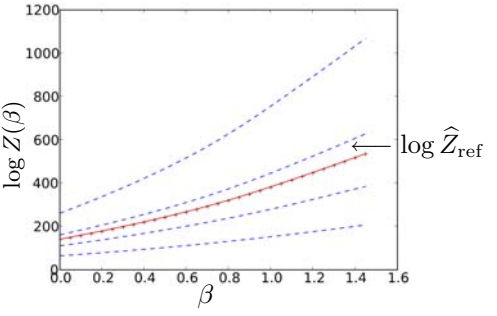


Fig. 1. The blue curves correspond to path-sampling estimates of the *reference* log-PFs $\log \widehat{Z}_{\mathcal{G}_p}(\beta_k), p = 1:4$. The red curve defines the ground truth (path-sampled $\log \widehat{Z}_{\mathcal{T}}$). Our extrapolation method provides the crossed-line ($\times$–) log-PF estimate $\log \widetilde{Z}_{\mathcal{T}}$.

Here, $(c_{\mathcal{T}}, c_{\mathcal{G}_p})$ and $(s_{\mathcal{T}}, s_{\mathcal{G}_p})$ are the number of cliques and sites of the Ising fields defined over $\mathcal{T}$ and $\mathcal{G}_p$. $\widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_p)$ corresponds to the $Z_{\mathcal{T}}(\beta)$ estimate computed using $\mathcal{G}_p$. Our extrapolation formula (3) is built up according to two principles directly derived from the PF definition: *i.)* an unbiased asymptotic approximation error ($\lim_{\beta \to +\infty} \mathcal{A}_{\mathcal{T}}(\beta, \mathcal{G}_p) = 0$) and *ii.)* an exact approximation of the first derivative of $\log Z_{\mathcal{T}}(\beta)$ for $\beta \to 0^+$. Interestingly, our extrapolation technique makes also possible the analytical computation of the approximation error $\mathcal{A}(\beta, \mathcal{G}_p)$ at $\beta = 0$, a value for which the error is maximal when $\mathcal{L}_{\mathcal{T}}(\mathcal{G}_p)$ is sufficiently low. This property were empirically verified on more than 300 fields simulated over regular and non-regular grids. The reference grid is then exhibited using a min-max principle, which consists in minimizing the error $\mathcal{A}(0, \mathcal{G}_p)$ wrt all reference grids $(\mathcal{G}_p)_{p=1:P}$ when the homogeneity of $\mathcal{T}$ and $\mathcal{G}_p$ is similar:

$$\mathcal{G}_{\text{ref}} = \underset{(\mathcal{G}_p)_{p=1:P}}{\arg\min}\ \mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) \quad \text{subject to} \quad \mathcal{L}_{\mathcal{T}}(\mathcal{G}_p) \leq \epsilon \tag{4}$$

$$\text{with} \quad \mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p) \triangleq \|(s_{\mathcal{T}} - 1) - c_{\mathcal{T}}(s_{\mathcal{G}_p} - 1)/c_{\mathcal{G}_p}\|^2 / s_{\mathcal{T}}^2 \tag{5}$$

where ϵ is positive threshold fixed by hand[3]. In practice, we first compute $\mathcal{L}_{\mathcal{T}}(\mathcal{G}_p)$ and $\mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p)$, $\forall \mathcal{G}_p$ and only keep the subset $\mathcal{S}$ of reference grids for which the constraint $\mathcal{L}_{\mathcal{T}}(\mathcal{G}_p) \leq \epsilon$ is fulfilled and $\mathcal{A}_{\mathcal{T}}(0, \mathcal{G}_p)$ is below another threshold (typically 4%). If $\mathcal{S}$ is empty, $\log Z_{\mathcal{T}}(\beta)$ is estimated using path-sampling. Otherwise, in a second stage, $\mathcal{G}_{\text{ref}}$ is exhibited from $\mathcal{S}$ as the minimizer of $\mathcal{A}_{\mathcal{T}}(0, \mathcal{S})$ and the log-PF estimate in $\mathcal{T}$ is thus given by $\log \widetilde{Z}_{\mathcal{T}}(\beta, \mathcal{G}_{\text{ref}})$ according to Eq. (3).

Our method is illustrated in Fig. 1 with $P = 4$: by comparing the distance between the blue and red curves at $\beta = 0$, it appears that $\log \widehat{Z}_{\text{ref}}$ is the closest curve *above* the path-sampled red curve $\log \widehat{Z}_{\mathcal{T}}$ considered here as the ground truth. As shown in Fig. 1, our extrapolation $\log \widetilde{Z}_{\mathcal{T}}$ is superimposed on $\log \widehat{Z}_{\mathcal{T}}$.

[3] We used $\epsilon = 0.02$ in our experiments.

4 Results

4.1 Simulated Ising Fields

For validation purpose, we compared log-PF estimates computed using our extrapolation technique with the linear and bilinear alternatives proposed in [8]. Here, the ground truth was given by the PF estimates computed using path-sampling. *Reference* and *test* grids are either *regular* or *irregular*. Each subset of reference or test grids contains at least 30 grids whose number of voxels varied between 5^3 and 16^3. Irregular grids were stochastically generated as a function of a regularization level α taking its values between 0.2 (highly irregular) and 0.5 (almost regular). All tests were performed on 3D Ising fields defined using a 6-connectivity system. Percentages of approximation errors are shown in Table 1.

Table 1. Mean maximal approximation errors (given in %) over regular and irregular test grids reported for the linear, bilinear and extrapolation techniques

Test grid		Scheme / Reference grid E=Extrapolation, B=bilinear, L=linear / R=regular, I=irregular				
		E / I & R	B / I	B / R	L / I	L / R
regular	small	**0.639**	3.84	66.3	93.0	2728
	medium	2.77	**0.991**	2.17	6.37	49.5
	large	3.68	**1.31**	2.48	7.18	19.4
irregular	$\alpha = 0.2$	**0.375**	1.29	94.6	83.9	3270
	$\alpha = 0.3$	**0.281**	0.784	2.91	18.3	219
	$\alpha = 0.4$	0.621	**0.264**	3.23	8.28	34.8
	$\alpha = 0.5$	**0.693**	1.27	1.96	1.52	34.2

The bilinear and extrapolation methods clearly outperform the linear one. Moreover, as shown in col. (B/ R) and rows (regular small and irregular, $\alpha = 0.2$), the bilinear method provides inaccurate estimates when there are strong topological differences between the *reference* and *test* grids. The regular reference grids are actually composed of large grids with cubic, planar and curvilinar shapes whereas those lying in (regular small) and (irregular $\alpha = 0.2$) are very small and highly sparse. In this case, our extrapolation method detects such differences and still succeeds in providing reliable log-PF estimates. While the linear/bilinear methods take all *reference* grids into consideration to derive a log-PF approximation, our approach computes a log-PF estimate using the most appropriate *reference* grid. Hence, the larger the set of reference grids the more accurate our extrapolation method becomes. This explains why the *reference* subsets are successfully mixed in the proposed approach, as shown in the first column of Table 1. Interestingly, when both the reference and test grids are non-regular the bilinear method may provide a competitive alternative to our extrapolation technique.

4.2 Real fMRI Dataset

Our extrapolation algorithm was applied to the spatially adaptative regularization of real fMRI data recorded during an event-related experiment designed to

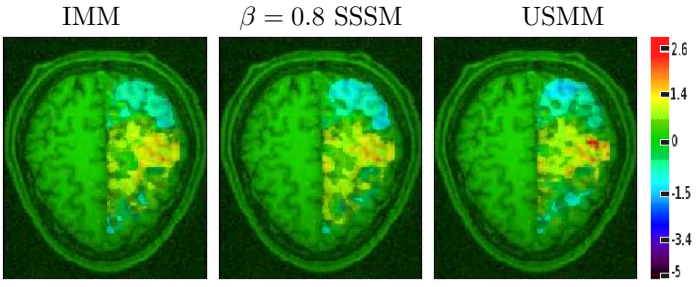


Fig. 2. From left to right: comparison of the IMM, SSMM and USMM models wrt the estimated normalized constrat maps: left vs. right auditory clicks: $\widehat{\boldsymbol{a}}^{\mathrm{LAC}} - \widehat{\boldsymbol{a}}^{\mathrm{RAC}}$

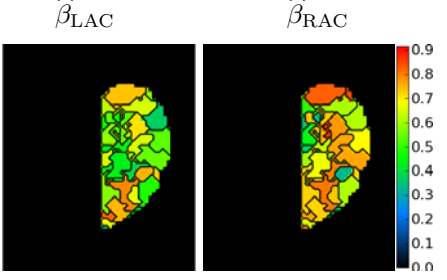


Fig. 3. Left: $\widehat{\beta}_{\mathrm{LAC}}$ parcel-dependent map computed for the LAC condition. **Right**: $\widehat{\beta}_{\mathrm{RAC}}$ parcel-dependent map computed for the RAC condition.

quickly map main sensory cortices (auditory, visual, motor) as well as higher cognitive functions (reading, computation). Acquisition consisted of a single session (125 scans, TR=2.4 s, 64x64x32 volumes). The paradigm comprised sixty stimuli, declined in 10 experimental conditions.

We compare three versions of the JDE procedure: Independent Mixture Models (IMM), Supervised SMM (SSMM, $\beta = 0.8$) and unsupervised SMM (USMM), in order to assess the impact of the adaptive spatial correlation model. Fig. 2 shows normalized contrast maps ($\widehat{\boldsymbol{a}}^{\mathrm{LAC}} - \widehat{\boldsymbol{a}}^{\mathrm{RAC}}$) of auditory induced left versus right clicks (LAC vs RAS). As expected, the activations lie in the contralateral right motor cortex. Here, only USMM is more sensitive illustrating therefore the advantage of an *adaptive* spatial correlation model. Indeed, estimated $\widehat{\beta}^{\mathrm{PM}}$ with USMM for the left auditory click was 0.56 so that the supervised setting of SSMM with $\beta = 0.8$ leads to too much correlation and less sensitive results.

Interestingly, Fig. 3 depicts the parcel-dependent maps of the $\widehat{\beta}^{\mathrm{PM}}$ estimates for the RAC and LAC experimental conditions. The gain in sensitivity in the USMM contrast map ($\widehat{\boldsymbol{a}}^{\mathrm{LAC}} - \widehat{\boldsymbol{a}}^{\mathrm{RAC}}$) clearly results from a difference in the amount of spatial regularization introduced between the two conditions involved in the contrast. A lower regularization level is estimated ($\widehat{\beta}_{\mathrm{LAC}} \approx 0.5$ vs. $\widehat{\beta}_{\mathrm{RAC}} \approx 0.75$) in parcels located in the right motor cortex since the BOLD signal is

stronger for the LAC condition than for the RAC one in these regions. Hence, the noise induced by the RAC condition is smoothed using a large amount of spatial regularization.

On these real fMRI data, our extrapolation scheme provides log-PFs estimate for a brain parcellation $(\mathcal{P}_\gamma)_{\gamma=1:300}$ and $(\mathcal{G}_p)_{p=1:50}$ reference grids. In terms of computational complexity, these log-PF estimates were computed in about ten seconds, a very appealing approach in comparison to path-sampling, which requires about one hour for estimating all log-PF estimates for a negligable gain in accuracy (less than 3%). Finally, we did not observe any significant difference between the USSM effect maps derived using path sampling and our extrapolation scheme (results not shown).

5 Discussion and Conclusion

In order to make spatially adaptive regularization feasible, the considered joint detection-estimation of brain activity from *unsmoothed* fMRI data requires a reliable and fast estimation of 3D Ising field partition function. To this end, an extrapolation algorithm that exploits pre-computed path-sampled log-PF estimates on reference grids has been proposed. The approximation error is controlled so that the approach defaults to the robust path-sampled PF estimates if no suitable reference candidate is found. Obviously, efficiency is conditionned by the number of reference grids, and more importantly by their similarity to the topologies encountered in the conducted analysis. In practice, about ten additional problem-specific reference grids are enough to provide good PF estimates.

Using our fast extrapolation technique, the computational burden remains acceptable since whole brain data analysis at the subject level takes about 1h30. The application to real fMRI data showed a gain in statistical sensitivity for the unsupervised version. In order to be properly validated, these promising within-subject results have to be confirmed in a group-level analysis.

References

1. Vincent, T., Ciuciu, P., Idier, J.: Spatial mixture modelling for the joint detection-estimation of brain activity in fMRI. In: 32th Proc. IEEE ICASSP, Honolulu, Hawaii, vol. I, pp. 325–328 (2007)
2. Meng, X., Wong, W.: Simulating ratios of normalizing constants via a simple identity: a theoretical exploration. Statistica Sinica 6, 831–860 (1996)
3. Gelman, A., Meng, X.L.: Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. Statistical Science 13, 163–185 (1998)
4. Fernández, S., Green, P.J.: Modelling spatially correlated data via mixtures: a Bayesian approach. J. R. Statist. Soc. B 64, 805–826 (2002)
5. Woolrich, M., Behrens, T., Beckmann, C., Smith, S.: Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data. IEEE Trans. Med. Imag. 24, 1–11 (2005)

6. Makni, S., Idier, J., Vincent, T., Thirion, B., Dehaene-Lambertz, G., Ciuciu, P.: A fully Bayesian approach to the parcel-based detection-estimation of brain activity in fMRI. Neuroimage 41, 941–969 (2008)
7. Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.B.: Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. Hum. Brain Mapp. 27, 678–693 (2006)
8. Trillon, A., Idier, J., Peureux, P.: Unsupervised Bayesian 3D reconstruction for non-destructive evaluation using gammagraphy. In: EUSIPCO, Suisse (2008)

Parcellation of fMRI Datasets with ICA and PLS-A Data Driven Approach

Yongnan Ji[1], Pierre-Yves Hervé[2], Uwe Aickelin[1], and Alain Pitiot[2]

[1] School of Computer Science, University of Nottingham,
Jubilee Campus, Wollaton Road, Nottingham, NG8 1BB, U.K.
yxj@cs.nott.ac.uk
[2] Brain and Body Centre, University of Nottingham,
University Park, Nottingham, NG7 2RD, U.K.

Abstract. Inter-subject parcellation of functional Magnetic Resonance Imaging (fMRI) data based on a standard General Linear Model (GLM) and spectral clustering was recently proposed as a means to alleviate the issues associated with spatial normalization in fMRI. However, for all its appeal, a GLM-based parcellation approach introduces its own biases, in the form of a priori knowledge about the shape of Hemodynamic Response Function (HRF) and task-related signal changes, or about the subject behaviour during the task.

In this paper, we introduce a data-driven version of the spectral clustering parcellation, based on Independent Component Analysis (ICA) and Partial Least Squares (PLS) instead of the GLM. First, a number of independent components are automatically selected. Seed voxels are then obtained from the associated ICA maps and we compute the PLS latent variables between the fMRI signal of the seed voxels (which covers regional variations of the HRF) and the principal components of the signal across all voxels. Finally, we parcellate all subjects data with a spectral clustering of the PLS latent variables.

We present results of the application of the proposed method on both single-subject and multi-subject fMRI datasets. Preliminary experimental results, evaluated with intra-parcel variance of GLM t-values and PLS derived t-values, indicate that this data-driven approach offers improvement in terms of parcellation accuracy over GLM based techniques.

1 Introduction

Inter-subject parcellation based on a standard General Linear Model (GLM) and spectral clustering was recently proposed as a means to alleviate the issues associated with spatial normalization in the analysis functional Magnetic Resonance Imaging (fMRI) datasets: lack of true anatomical correspondence, inaccuracy of the normalization process (see [1] for an in-depth overview), etc. In a parcellation framework, voxels are first clustered into functionally homogeneous regions or *parcels*. Then, the parcellations are then homogenised across subjects, so that statistics can be carried out at the parcel level rather than at the voxel level. Here we focus on the optimization of the first step of the parcellation scheme.

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 984–991, 2009.

We present a data-driven, model-free, parcellation technique, based on Independent Component Analysis (ICA) and Partial Least Squares (PLS, [2]) instead of a GLM, so as to use more of the information contained within the fMRI time series. First, a number of independent components are automatically selected. Seed voxels are then obtained from the associated ICA maps and we compute the PLS latent variables between the fMRI signal of the seed voxels (which covers regional variations of the stimuli related BOLD responses) and the principal components of the signal across all voxels. Finally, we parcellate all subjects data with a spectral clustering of the PLS latent variables. We also introduce PLS t-values as an alternative way to validate parcellation results.

We detail our approach in the following Section 2. Preliminary results are given in Section 3, where we also compare them to GLM-based parcellation.

2 Materials and Methods

2.1 Datasets and Preprocessing

We applied our method to two functional datasets: a single-subject fMRI experiment with a standard finger tapping task and a multi-subject experiment where volunteers were presented with hand gestures or face expressions [3].

Single-subject data were acquired on a Philips Intera 1.5T with a TR of 3s and a sequential finger tapping task auditorily paced with a metronome. The auditory signals were given every 0.6 seconds. The digit order of the tapping was 1 - 3 - 2 - 4, repeated 6 times in each period, with a 14.4 second rest between periods. The period of one on-and-off block was then 28.8 seconds.

Multi-subject data, our main concern in this paper, were acquired from 25 subjects viewing angry gestures or expressions. Scanning was performed on a Philips Intera 1.5T, with TR=3s. During the scan, four types of visual stimuli are given to the subjects, which are angry hand gestures, neural hand gestures, angry facial expression and neural facial expression.

Both datasets were preprocessed with FSL for slice-timing, motion correction and registration [4,5].

2.2 Independent Components Selection

For the multi-subject experiment, we used FSL to decompose the input fMRI data into independent components (ICs). We obtained between 30 and 60 ICs per subject, for a total of 1203 ICs. Here we propose to use a hierarchical clustering approach, similar in spirit to Partner Matching [6] as a means to find the ICs that best capture the Blood Oxygen Level Dependent Haemodynamic (BOLD) response to the stimuli. This method is based on the assumption that very few of the 1203 ICs contain information about the stimuli-related BOLD responses. Consequently, the task-related ICs should be more similar with each other than with the other ICs since they share the same source. We aim to group those ICs that correspond to the response to the same task features in different subjects

together into one cluster. The other ICs which do not contain relevant (i.e. task-related) information should be grouped inside another cluster.

We take those constraints into account when design the similarity function to be used in our hierarchical clustering.

Let N_a and N_b be the number of ICs for subjects A and B respectively, with $\mathbf{IC}_i^{\mathrm{A}}$ and $\mathbf{IC}_j^{\mathrm{B}}$ the ith IC of subject A and jth IC of subject B. $t = 1, \ldots, T$ is the time index. Their correlation coefficients is given by:

$$\rho(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}}) = \frac{\sum_{t=1}^{T}(\mathbf{IC}_i^{\mathrm{A}}(t) - \overline{\mathbf{IC}_i^{\mathrm{A}}})(\mathbf{IC}_i^{\mathrm{B}}(t) - \overline{\mathbf{IC}_i^{\mathrm{B}}})}{\sqrt{\sum_{t=1}^{T}(\mathbf{IC}_i^{\mathrm{A}}(t) - \overline{\mathbf{IC}_i^{\mathrm{A}}})^2}\sqrt{\sum_{t=1}^{T}(\mathbf{IC}_i^{\mathrm{B}}(t) - \overline{\mathbf{IC}_i^{\mathrm{B}}})^2}}, \quad (1)$$

The normalized correlation coefficients ρ_{norm} is:

$$\rho_{\mathrm{norm}}(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}}) = \frac{\rho(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}}) - \mathrm{mean}(\rho(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}})|_{j=1,2,\ldots,N_b})}{\mathrm{std}(\rho(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}})|_{j=1,2,\ldots,N_b})} \quad (2)$$

Since the aim of the clustering is to put similar ICs from different subjects into one cluster, all the ICs of the same cluster should come from different subjects, therefore we need to set the similarity between ICs of the same subject to 0. The similarity between two ICs is finally defined as

$$S(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}}) = \begin{cases} 0 & \text{if } A = B \\ \min(\rho_{\mathrm{norm}}(\mathbf{IC}_i^{\mathrm{A}}, \mathbf{IC}_j^{\mathrm{B}}), \rho_{\mathrm{norm}}(\mathbf{IC}_j^{\mathrm{B}}, \mathbf{IC}_i^{\mathrm{A}})) & \text{other wise .} \end{cases} \quad (3)$$

In the case of the single-subject data, the ICs representing BOLD signals could not be selected by comparing ICs across subjects as above. Therefore, we manually picked those ICs that best matches the canonical HRF-convoluted task design from the 34 ICs produced by FSL.

2.3 Seed Selection

In order to calculate the PLS latent variables that best capture the BOLD response, a number of seeds representing different active regions should be selected. For instance, in a GLM-based parcellation approach, we could select as seeds the voxels with the largest t-values. Here, we pick them on the basis of the ICA results.

Within each IC map [5], the first seed is chosen as the voxel with the largest value. The second seed is then chosen, amongst the voxels at least R voxels away from the first seed voxel, as the voxel with the largest IC map value. The iterative process is repeated until all the seeds have been selected.

In the multi-subject case, two IC maps were used. We picked $R = 6$ voxels and obtained $N_{seed} = 15$ seeds for each map. In the single-subject data, 30 seeds were selected from each IC map with $R = 6$.

2.4 PCA/PLS Feature Space for Parcellation

Let $\mathbf{X}_{V \times T} = x_{ij}$ denote the data matrix, where each row corresponds to the fMRI signal of a given voxel. Then, we propose to denoise the signal with PCA.

We first center the signal at each voxel by substracting its mean, before decomposing $\mathbf{X}_{V\times T}$ into $\mathbf{P}_{\mathrm{PCA}}$ and $\mathbf{T}_{\mathrm{PCA}}$ using Principal Component Analysis (PCA) as $\mathbf{X}_{V\times T} = \mathbf{P}_{\mathrm{PCA}}\mathbf{T}'_{\mathrm{PCA}}$ [7,8]. $\mathbf{T}'_{\mathrm{PCA}}$ is the transpose of the PCA score matrix of $\mathbf{X}$ (the matrix whose columns are the Principal Components (PCs) of the fMRI data), and $\mathbf{P}_{\mathrm{PCA}}$ is the PCA loading matrix. Ranking the PCs according to the variance they cover, the first few. PCs usually have exceptionally high variances. And the last PCs are slow-variant artefacts. These PCs are considered as noise and removed.

Let $\mathbf{D}_{T\times N_{seed}}$ represent the fMRI signals of the seed voxels, where each column correspond to the fMRI signal in a given seed. We then use the PCs in matrix $\mathbf{T}_{\mathrm{PCA}}$ for the prediction of $\mathbf{D}$ with Partial Least Square (PLS). The original design of PLS is to predict $\mathbf{D}$ with the components decomposed from $\mathbf{T}_{\mathrm{PCA}}$ and $\mathbf{D}$ as regressor. These components, the latent variables, should contain the information from both $\mathbf{T}_{\mathrm{PCA}}$ and $\mathbf{D}$. Here PLS is used to calculate the time series components that represent the individual specific functional activity signals. We decompose $\mathbf{T}_{\mathrm{PCA}}$ into the product of $\mathbf{T}_{\mathrm{PLS}}$ and $\mathbf{P}'_{\mathrm{PLS}}$ with $\mathbf{T}'_{\mathrm{PLS}}\mathbf{T}_{\mathrm{PLS}} = \mathbf{I}$. $\mathbf{D}$ is predicted as $\hat{\mathbf{D}} = \mathbf{T}_{\mathrm{PLS}}\mathbf{B}\mathbf{C}'$, where the columns of $\mathbf{T}_{\mathrm{PLS}}$, $\mathbf{t}_i, i = 1, 2, ..., K$, are the latent vectors of size $T \times 1$. $\mathbf{B}$ is a diagonal matrix with the "regression weights" as diagonal elements and $\mathbf{C}$ is the "weight matrix" of the dependent variables [9].

Given $\mathbf{T}_{\mathrm{PCA}}$ and $\mathbf{D}$, the latent vectors could be chosen in a lot of different ways. The canonical way is to find the latent vectors that maximize the covariance between the columns of $\mathbf{T}_{\mathrm{PLS}}$ and $\mathbf{D}$. Specifically, the first latent vector is calculated as $\mathbf{t}_1 = \mathbf{T}_{\mathrm{PCA}}\mathbf{w}_1$ and $\mathbf{u}_1 = \mathbf{D}\mathbf{c}_1$ with the constrains that $\mathbf{t}'_1\mathbf{t}_1 = 1$, $\mathbf{w}'_1\mathbf{w}_1 = 1$ and $\mathbf{t}'_1\mathbf{u}_1$ be maximal. Then first component is subtracted from $\mathbf{T}_{\mathrm{PCA}}$ and $\mathbf{D}$, and the rest latent variables are calculated iteratively as the above until $\mathbf{T}_{\mathrm{PCA}}$ becomes a null matrix. The first PLS latent variables are the signals of interest. Let $\mathbf{X}_0$ be derived from $\mathbf{X}$ after the signal variance has been revmoed: $\mathbf{x}_0 = \mathbf{x}/||\mathbf{x}||$, where $\mathbf{x}$ and $\mathbf{x}_0$ are the row vectors of $\mathbf{X}$ and $\mathbf{X}_0$. We use the covariances between fMRI signals and latent variables, $\mathbf{r}_i = \mathbf{X}_0\mathbf{t}_i$, as feature space for parcellation.

2.5 Parcellation Method

Following Thirion et al.[1], we chose spectral clustering for parcellation. This method represents the relationships between voxels as a graph whose vertices correspond to the voxels with the functional distance between voxels (GLM-based in their approach) associated to the edges. The complete distance matrix Δ_G between all pairs of voxels is obtained by integrating the local distances along the paths in the graph. Singular Value Decomposition (SVD) is applied to the centered square distance matrix $\overline{\Delta_G}$. Finally, they apply C-means clustering to the singular vectors with largest singular values. Please see [1] for details.

Here, rather than using GLM to define local distances, we use our r_i's:

$$d(v, w) = ||\mathbf{r}(v) - \mathbf{r}(w)|| = \sqrt{(\mathbf{r}(v) - \mathbf{r}(w))(\mathbf{r}(v) - \mathbf{r}(w))'}, \tag{4}$$

where v and w are two neighbouring voxels and $\mathbf{r}(v) = [\mathbf{r}_1(v)\ \mathbf{r}_2(v))... \mathbf{r}_K(v)]$.

3 Results

3.1 Intra-parcel Functional Homogeneity

As in [1], we use the intra-parcel functional variance to assess the quality of the parcellation. However, instead of using GLM parameters to represent the functional information, we use GLM t-values and PLS t-values (described below) for each regressor as functional features. Let Nr be the number of regressors and $\mathbf{f}^i \in R^{Nr\times 1}$ be the vector of t-values for voxel i. For any parcel p, the functional variance of p, $v(p)$, is:

$$v(p) = \sqrt{\sum_{k=1}^{Nr}(\mathrm{std}(f_k^i))^2}, \quad \text{where, } \mathbf{f}^i = [f_1^i \ f_2^i \ \ldots \ f_{Nr}^i]' \text{ and } i \in p \tag{5}$$

The distribution of $v(p)$ across all parcels is used to compare the accuracy of the parcellations.

PLS *t*-values. Given a design matrix $\mathbf{Y} \in R^{T\times Nr}$, where $\mathbf{y}_k \in R^{T\times 1}$ is the kth column of $\mathbf{Y}$, instead of using $\mathbf{D}$, the regressor $\mathbf{y}_k$ is used to calculate latent variables as in section 2.4. If $\mathbf{r}_k$ is the covariance between the fMRI time series and the first latent variable, then,

$$t = \frac{r\sqrt{T-2}}{1-r^2} \tag{6}$$

has a t-distribution with $T-2$ degrees of freedom. The null hypothesis of this test is that the signal of that voxel is not covariant with the PLS components. Thus, we can generate statistical maps to represent the significance of the covariance between the signals in each voxel and the first latent variable calculated from data and the kth regressor of the design matrix.

3.2 Results on Single-Subject Data

As mentioned above in section 2.2, our automatic IC selection approach cannot be applied to single subject data. Here, we manually selected the IC whose time course best matches the experiment design, to be used in the seed selection process. The fMRI signals of the whole brain are decomposed into PCs. It should be noted that the PCs covering the largest variance in $\mathbf{X}$ are not the most interesting signals in the fMRI dataset. Indeed it appears that the respiratory, cardiac or instrumental artifacts may have a larger influence on the BOLD signal than the task. Such noise-related PCs are removed before the PLS step.

In Fig. 1, the first two latent variables are shown. The shape of the first latent variable matches the experiment design. It can be considered as a subject-specific response model, which will allow a better detection of task-related activity. Based on the latent variables, the whole brain is parcellated into 600 parcels using a spectral clustering as explained in section 2.5. The intra-parcel variances of GLM

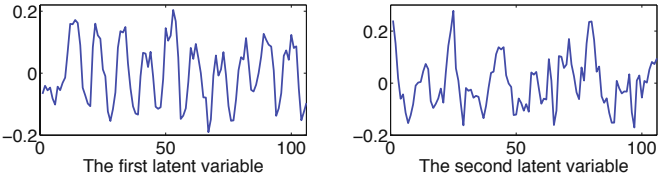


Fig. 1. First two PLS components

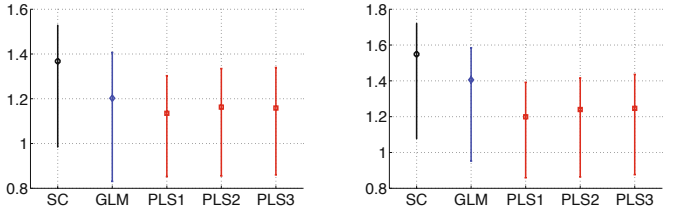


Fig. 2. Comparison of Parcellation results from Spatial Clustering (SC), Spectral Clustering with GLM (GLM), Spectral Clustering with 1 PLS latent variable (PLS1), 2 PLS latent variable (PLS2) and 3 PLS latent variable (PLS3)

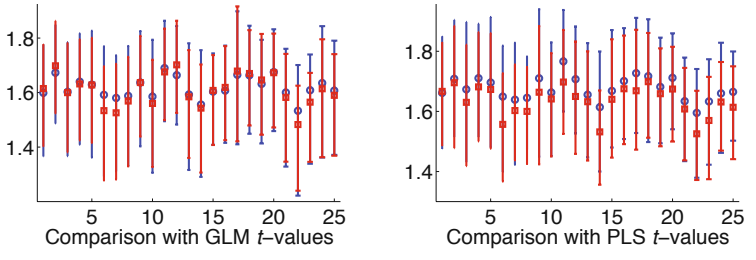


Fig. 3. Comparison of functional intra-parcel homogeneity

t-values and PLS t-values are used to compare parcellation results based on GLM and PLS. Here, the bars illustrate the mean, the first and the third quartile of the t-values variance of 600 parcels from each method. From Fig. 2, we can see that with both functional measures, spectral clustering with PLS increases the intra-parcel functional homogeneity. One latent variable is optimal for the parcellation of this dataset.

3.3 Results on Multiple-Subject Data

Using the similarity matrix described in section 2.3 and Ward's linkage, we grouped all the ICs into three clusters. The ICs in cluster 1 match the first and second task regressors. The ICs in cluster 2 match the third and fourth task regressor. Meanwhile, in feature space, these two clusters keep large distances from the rest of the ICs. There are 20 ICs in cluster 1 from 19 subjects and 20 ICs in cluster 2 from 20 subjects. For each subject we use the ICs from these for

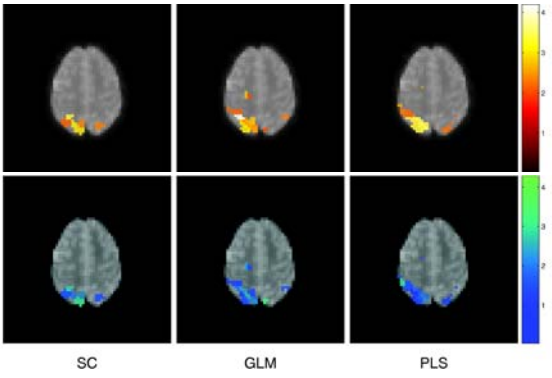


Fig. 4. Parcels respond to stimuli. Parcels with average GLM t-values larger than 2 are shown. The first row shows activation. The second row shows intra-parcel variance. Three columns show the results from three parcellation methods.

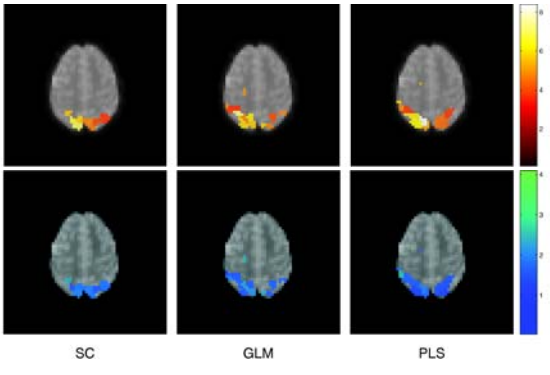


Fig. 5. Parcels respond to stimuli. Parcels with average PLS t-values larger than 3 are shown.

sampling seed voxels. If a subject doesn't have an IC in cluster 1 or 2, we use the ICs that are closest to those clusters to sample the seeds.

All the datasets are then parcellated into 600 parcels. The functional variances of GLM parcellation and ICA-PLS parcellation on 25 subjects are compared in Fig. 3. Subjects are split across the horizontal axis. The parcels respond to the stimuli of angry hand gestures are shown in Fig. 4. The activation and intra-parcel functional variance are evaluated with GLM t-values. Similarly, in Fig. 5, the parcellation results are evaluated with PLS t-values.

4 Conclusion and Future Work

We presented a data driven method for parcellation of fMRI data. Preliminary experimental results indicate that such approach adapts to the variability of the BOLD response across subjects and increase the accuracy of the parcellation.

The cost of this improvement is the complexity of parcellation. Future work will tackle the homogenisation of those parcellations across different subjects.

Acknowledgments. This research is funded by the European Commission FP6 Marie Curie Action Programme (MEST-CT-2005-021170).

References

1. Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.-B.: Dealing with the shortcomings of spatial normalization: multi-subject parcellation of fmri datasets. Human brain mapping 27, 678–693 (2006) PMID: 16281292
2. McIntosh, A.R., Lobaugh, N.J.: Partial least squares analysis of neuroimaging data: applications and advances. NeuroImage 23, S250–S263
3. Grosbras, M., Paus, T.: Brain networks involved in viewing angry hands or faces. Cereb. Cortex 16, 1087–1096 (2006)
4. Fmrib software library (August 2007)
5. Beckmann, C., Smith, S.: Probabilistic independent component analysis for functional magnetic resonance imaging (2002)
6. Wang, Z., Peterson, B.S.: Partner-matching for the automated identification of reproducible ICA components from fMRI datasets: algorithm and validation. Human Brain Mapping 29, 875–893 (2008) PMID: 18058813
7. Kerrouche, N., Herholz, K., Mielke, R., Holthoff, V., Baron, J.: 18FDG PET in vascular dementia: differentiation from alzheimer's disease using voxel-based multivariate analysis. Journal of Cerebral Blood Flow and Metabolism: Official Journal of the International Society of Cerebral Blood Flow and Metabolism 26, 1213–1221 (2006) PMID: 16525414
8. Zuendorf, G., Kerrouche, N., Herholz, K., Baron, J.: Efficient principal component analysis for multivariate 3D voxel-based mapping of brain functional imaging data sets as applied to FDG-PET and normal aging. Human Brain Mapping 18(1), 13–21 (2003)
9. Abdi, H.: Partial least squares regression (pls-regression). In: Lewis Beck, M., Bryman, A., Futing, T. (eds.) Encyclopedia for research methods for the. Social Sciences, vol. 2003, pp. 792–795. Sage, Thousand Oaks (2003)

Adjusting the Neuroimaging Statistical Inferences for Nonstationarity

Gholamreza Salimi-Khorshidi[1], Stephen M. Smith[1], and Thomas E. Nichols[1,2]

[1] Centre for Functional MRI of the Brain (FMRIB),
University of Oxford, Oxford, UK
reza@fmrib.ox.ac.uk
[2] GlaxoSmithKline Clinical Imaging Centre, London, UK

Abstract. In neuroimaging cluster-based inference has generally been found to be more powerful than voxel-wise inference [1]. However standard cluster-based methods assume stationarity (constant smoothness), while under nonstationarity clusters are larger in smooth regions just by chance, making false positive risk spatially variant. Hayasaka et al. [2] proposed a Random Field Theory (RFT) based nonstationarity adjustment for cluster inference and validated the method in terms of controlling the overall family-wise false positive rate. The RFT-based methods, however, have never been directly assessed in terms of homogeneity of local false positive risk. In this work we propose a new cluster size adjustment that accounts for local smoothness, based on local empirical cluster size distributions and a two-pass permutation method. We also propose a new approach to measure homogeneity of local false positive risk, and use this method to compare the RFT-based and our new empirical adjustment methods. We apply these techniques to both cluster-based and a related inference, threshold-free cluster enhancement (TFCE). Using simulated and real data we confirm the expected heterogeneity in false positive risk with unadjusted cluster inference but find that RFT-based adjustment does not fully eliminate heterogeneity; we also observe that our proposed empirical adjustment dramatically increases the homogeneity and TFCE inference is generally quite robust to nonstationarity.

1 Introduction

When detecting changes in functional or structural brain image data, it is necessary to have powerful inference methods that offer precise control of false positive risk. To assess the evidence of a change at each voxel of a statistic image, the two most common approaches are voxel- and cluster-based inferences. While voxel-wise methods use a single threshold to classify signals as *real*, cluster-based inference defines clusters as contiguous voxels whose intensity exceeds a predefined cluster-forming threshold u_c, and then detects signals based on the spatial extent of a cluster. Cluster-based inference is known to have a higher sensitivity compared to voxel-intensity-based tests when the signal is spatially extended.

Cluster-size tests have been widely used under different implementations [1,3], including simulation-based tests [4,5], random field theory (RFT-based) tests [6],

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 992–999, 2009.

and permutation tests [7,8]. However most of these proceedures are based on a stationarity assumption, that the spatial autocorrelation function is shift-invariant. When stationarity assumption is violated, the sensitivity and specificity of the test depend on local smoothness of the image, which justifies the use of adjusted cluster size which measures cluster size relative to local smoothness [6,2]. Independent of the stationarity issue, cluster-based inference also is limited by the arbitrariness of its important u_c parameter and the amount of pre-smoothing. To address these problems, threshold-free cluster enhancement (TFCE) was introduced [9], which in essence integrates out the u_c parameter while produces an image of local cluster-like evidence of a signal and was shown to generally have better detection power while being less sensitive to the amount of smoothing used [9,10].

In this work we propose a new adjustment for nonstationarity based on the local empirical distribution of cluster size in a two-pass permutation method. We evaluate this new approach, in the context of both standard cluster-based and TFCE inferences. We compare the impact of using no adjustment, RFT-based adjustment, and our proposed empirical adjustment under various simulated and real data with spatially-varying smoothness.

2 Method

A fitted general linear model at a voxel i has residuals

$$\hat{\epsilon}_i = Y_i - X\hat{\beta}_i \tag{1}$$

where Y_i is the observed intensity (M×1), $\hat{\beta}_v$ is the parameter vector (P×1), X is the design matrix (M×P), and $\hat{\epsilon}_v$ is the residual error (M×1), the estimated residuals of the linear model fit ($\hat{\epsilon}$) can be used to yield a smoothness estimation. This estimator explains the spatial correlation structure of the SPM (statistical parameter map) by assuming that it can be modeled as being due to a convolution of the signal with a Gaussian filter with an unknown, but determinable, width (σ), which is used to estimate the effective resolution of the data.

2.1 Kiebel et al.'s Method

Using the RFT concepts [11], an unbiased estimator for the covariance of the partial derivatives (at direction j) at voxel i in a D-dimensional Gaussian random field is calculated as

$$\lambda_{i,j} = \frac{\nu - 2}{\nu - 1} \cdot \frac{1}{M} \sum_{t=1}^{M} \left(\frac{\partial S_{it}}{\partial x_j} \right)^2 \tag{2}$$

where S is the standardized error, ν is the number of degrees of freedom, and M is the number of observations (time points or subjects), which yields a voxel-wise estimate of smoothness as

$$RESEL_i = \prod_{j=1}^{D} \left(8 \cdot \ln(2)\right)^{1/2} \sigma_{i,j} \tag{3}$$

where $\sigma_{i,j} = (2\lambda_{i,j})^{-1/2}$, and $RESEL_i$ is the volume of a resolution element (resel) at voxel i (note that $1/RESEL$ is resels per voxel or RPV). As an alternative, a more robust estimate can be defined using just the control group observations or using all the observations after excluding any outlier (in terms of their smoothness estimates) observations.

2.2 Jenkinson et al.'s Method

As the smoothness extent decreases Kiebel's estimator becomes increasingly inaccurate. An alternative estimate can use the autocorrelation of the standard error at voxel i (S_i^2) and its cross-correlation with neighboring voxels (i.e., $SS_{i,j}$ for the next voxel in direction j) with a Gaussian autocorrelation function assumption [12,13]. A voxel-wise smoothness can be estimated using the σ from

$$\sigma_{i,j}^2 = \left(4 \cdot \ln\left(\frac{S_i^2}{SS_{i,j}}\right)\right)^{-1} \tag{4}$$

with the rest of the operation as before (which can also result in a robust estimate as in Kiebel's).

2.3 Empirical Cluster-Size Normalization

In FSL's *randomise* analysis[1], after N_P permutations, let $N_v \leq N_P$ be the number of permutations that clusters with sizes $S_1(v)$, $S_2(v)$, ... $S_{N_v}(v)$ hit the voxel v. The empirical cluster size per voxel (ECSPV) for this voxel is calculated as

$$ECSPV(v) = \left(\frac{\sum_{i=1}^{N_v} S_i(v)^E}{N_v}\right)^{1/E} \tag{5}$$

where E is the cluster size histogram's normalization parameter. Having ECSPV from the first run, the adjustment in the second run can either be voxel-wise

$$S_C^{vn} = \sum_{v \in C} \frac{1}{ECSPV(v)} \tag{6}$$

or cluster-wise

$$S_C^{cn} = \frac{S_C}{\sum_{v \in C} ECSPV(v)}. \tag{7}$$

Cluster-based inference using these normalized statistics (S_C^{vn} or S_C^{cn}) is expected to be adjusted for nonstationarity (estimating the smoothness/roughness of each area by using the cluster sizes hitting each voxel at different permutations under the null hypothesis).

[1] see http://www.fmrib.ox.ac.uk/fsl/randomise/

2.4 Empirical TFCE Normalization

TFCE is a tool developed based on the idea of cluster size accumulation on a range of possible cluster-forming thresholds [9]. In order to adjust the TFCE statistic for nonstationarity, either its corresponding cluster sizes can be adjusted or TFCE scores at each voxel can be empirically normalized by

$$ETPV(v) = \frac{\sum_{i=1}^{N_v} TFCE(v)}{N_v} \qquad (8)$$
$$TFCE^N(v) = \frac{TFCE(v)}{ETPV(v)}$$

where N_v is the number of permutations a voxel has a nonzero TFCE score and ETPV is the empirical TFCE per voxel.

2.5 Nonstationarity Assessment

Using null data, spatial variation in cluster-related inference's false positive rate (P-value) is used to assess different methods' performance in correction for nonstationarity. In case of using a stationary null data for a statistical inference, the output P-value volume should follow a uniform distribution ($U(0,1)$ with the mean of 0.5) at each voxel. Since $-\ln$ P-values are easier to visualize, note that if X has a uniform distribution, $-\ln(X)$ has an exponential distribution with parameter $\lambda = 1$, so $E[-\ln(X)] = \frac{1}{\lambda} = 1$, and $Var[-\ln(X)] = \frac{1}{\lambda^2} = 1$, and hence $E[-\log_{10}(X)] = \frac{1}{\ln(10)} = 0.4343$, and $Var[-\log_{10}(X)] = 0.4343$.

Thus, the deviation of an inference from this expected distribution can be a good indicator of the existence of nonstationarity in the image. Three statistical indicators (mean, standard deviation or SD, and coefficient of variation or CV) are employed to assess this deviation for each adjustment technique. To implement this idea, a group of permutations in the *first run* result in a distribution of cluster-related statistic. Then, for the same group of permutations in the *second run*, the resulting cluster-related statistic image is converted to a P-value volume with respect to the distribution from the first run. This P-value volume is then converted to $-\log_{10}(P)$ volume. Averaging these -$\log_{10}(P)$ volumes across all the permutations results in a single -$\log_{10}(P)$ volume whose mean, SD and CV indicates the extent of nonstationarity in the data. Note that, as a result of the averaging of $-\log_{10}$(P) images, the mean can be compared with the expected mean (0.4343), however the SD should be smaller than the expected SD (0.4343).

2.6 Data

To assess different methods' performance, both simulated and real data are tested. For the stationary null data simulation, two groups of 150×150×150 Gaussian noise ($\sim N(0,1)$) images (with 20 images in each group) are generated and smoothed with a Gaussian smoothing kernel (with σ=2, 3, 4 and 5 voxels). To avoid the nonstationarity at the edge, the outer 30 voxels are excluded. To

simulate the nonstationary data, two groups of 150×150×150 white noise images (with 20 images in each group) are first smoothed with three different 3D Gaussian kernels (representing the low, medium, and high smoothness extents). These images are combined in a way that an outer layer smoothed with σ_1 encloses a middle layer smoothed with σ_2, which encircles a core smoothed with σ_3 (referred to as $\sigma_1\sigma_2\sigma_3$). The center core is 30×30×30 voxels, centered within a 60×60×60 voxel middle layer, which itself was centered in a 90×90×90 volume. The combined image is smoothed again with a 3D Gaussian filter with σ =1.5 voxels (to eliminate discontinuities at the borders), and just as in the stationary case, to avoid the nonstationarity at the edges, outer 30 voxel are excluded.

To assess each method's performance on real data, null fMRI and VBM datasets are also fed into the analysis. The fMRI dataset is a pain study with 16 healthy subjects. The first-level analysis of the data using FSL includes motion correction and spatial smoothing (FWHM (full width half maximum) = 2, 4, 5, 7 and 10mm) prior to temporal model fitting (including autocorrelation correction), and a two-step registration to the MNI152 standard brain space. The null VBM dataset includes structural gray matter images of 35 healthy control subjects smoothed with different Gaussian kernel sizes (σ =2, 3, 4 and 5mm). Dividing these subjects into two groups results in a null data analysis (with no expected difference). The final assessment of the methods' performance is on real VBM data with three groups of subjects: 46 controls, 50 Alzheimer's disease (AD) and 57 mild cognitive impairment (MCI).

3 Results

Fig. 1 illustrates statistical indicators (mean, SD and CV of $-\log_{10}(P)$ volumes) of the previously mentioned analyses on both simulated and real data. Kiebel's and Jenkinson's method have very similar results, which is why only Jenkinson's results (referred to as RPV1) are shown (RPV0 refers to no adjustment). Also, results for $E = 1$ and 2 are not shown as $E = 2/3$ shows a better adjustment; *vw* and *cw* refer to voxel-wise and cluster-wise normalization. Fig. 2 and 3 illustrate the effect of different adjustment techniques in localizing the differences between two groups of subjects in real data, by assessing the change in P-value as a function of clusters' unadjusted P-value and average FWHM (see figure captions for details).

4 Discussion

Using cluster-size adjustment techniques, cluster sizes depend on both their connected component's size and the local roughness of their region. The results show that there is a substantial spatial variation in cluster-based inference's FPR that cannot be completely corrected. Using the uniformity indices, adjustments improve the homogeneity of the result image. According to the results, empirical adjustment causes a reduction of sensitivity, which may also be an indication of increased power of the inference. The optimal empirical adjustment is seen at

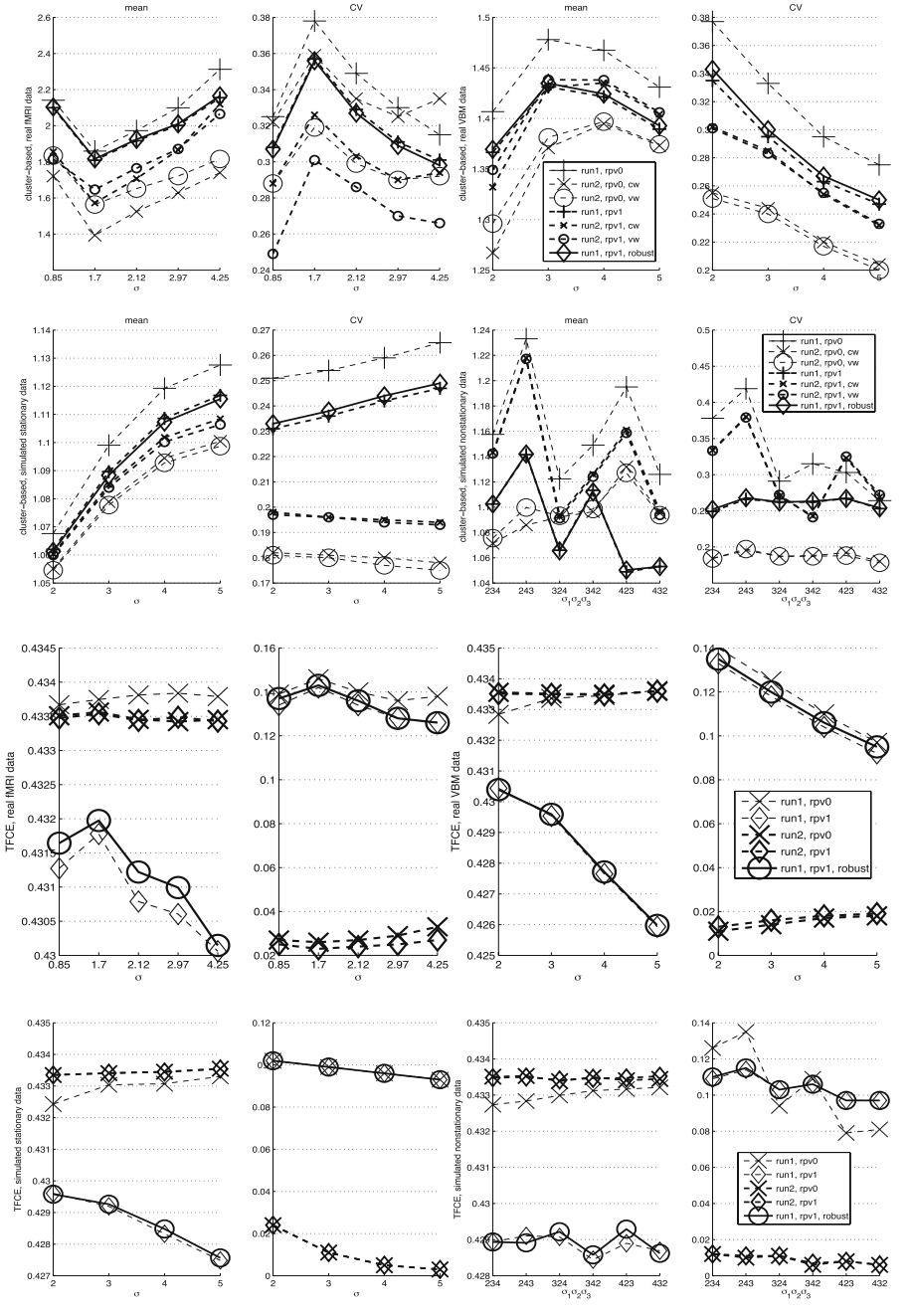


Fig. 1. The spatial mean and CV of -$\log_{10}(p)$ as a function of smoothing extent. These results show how adjustment can improve the stationarity of the cluster-related inferences and TFCE's robustness to nonstationarity in the data. Note that the legend in the first (third) row can be used for the third (fourth) row.

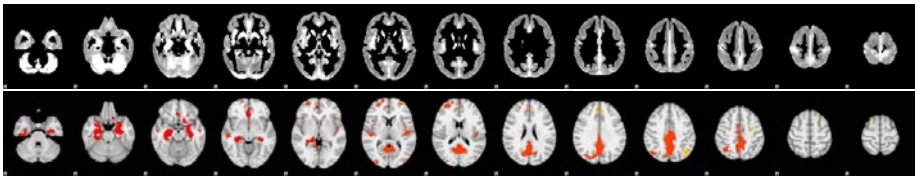


Fig. 2. Smoothness map corresponding the three-group VBM data (top image) and the resulting clusters after thresholding the T-statistic map at $T = 4$ (for the MCI-AD contrast). Smoothness estimates and clusters formed in this images are the basis of the analyses and the results sin Fig. 3. Note that the displayed slices are selected from z=-32mm to z=64mm including every eighth millimeters (in MNI coordinates).

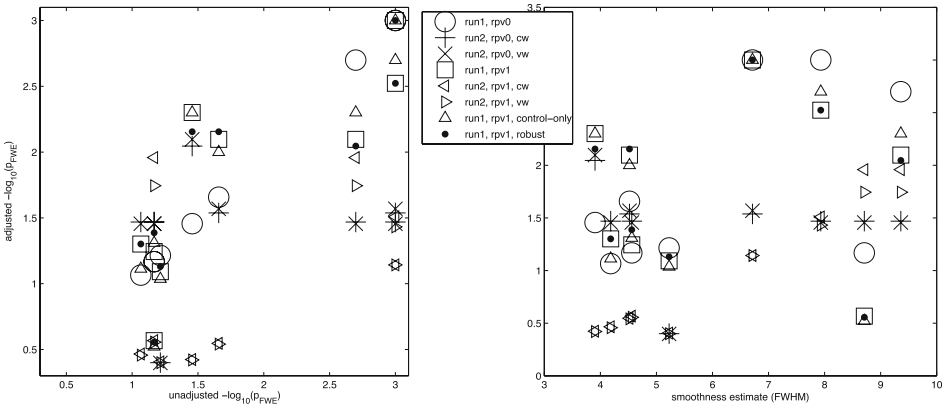


Fig. 3. The effect of the adjustment on significance level of clusters as a function of its FWHM (right column) and unadjusted P-value (left column). This figure illustrates the increase in sensitivity (specificity) in rough (smooth) regions after appropriate adjustments. Note that the clusters in this figure are shown in Fig. 2, which will remain the same for all the adjustment analyses. According to the right column, the expected variation of the significance (increase/decrease in rough/smooth regions) can be observed in empirical adjustment when no RFT-based adjustment is present, and RFT-based adjustment without empirical adjustment. Also, according to the left column, adjusting the cluster-based inference for nonstationarity seems to reduce the significance-level of the clusters, which agrees with the result in the top row to some extent (as more clusters are to be formed in smooth regions, where adjustment is expected to result in a significance decrease.

$E = 2/3$ and voxel-wise normalization in ECSPV calculation. The empirical adjustment is not recommended for adjusted cluster-sizes as the second correction applied to a unified field, may be similar to using $CS = \sum_{v \in C} 1$, which is the classic unadjusted cluster size. Note that, the use of cluster sizes to extract a voxel-wise characteristic (i.e., ECSPV) can be an imprecise estimate, because of the censoring (the P-value corresponding to a cluster, is not a precise voxel-wise

P-value). On the other hand, TFCE inference is very robust with respect to spatial variations in image smoothness. In both adjusted and unadjusted TFCE inferences, the summary measure of performance, perfectly matches the expected measure of a uniform image at all of the tested smoothing extents.

References

1. Friston, K., Holmes, A., Poline, J., Price, C., Frith, C.: Detecting activations in pet and fmri: levels of inference and power. NeuroImage 4, 223–235 (1996)
2. Hayasaka, S., Phan, K., Liberzon, I., Worsley, K., Nichols, T.: Nonstationary cluster-size inference with random field and permutation methods. Neuroimage 22(2), 676–687 (2004)
3. Poline, J., Worsley, K., Evans, A., Friston, K.: Combining spatial extent and peak intensity to test for activations in functional imaging. Neuroimage 5(2), 83–96 (1997)
4. Forman, S., Cohen, J., Fitzgerald, J., Eddy, W., Mintun, M., Noll, D.: Improved assessment of significant activation in functional magnetic resonance imaging (fmri): use of a cluster-size threshold. Magn. Reson. Med. 33, 636–647 (1995)
5. Ledberg, A., Åkerman, S., Roland, P.: Estimation of the probability of 3d clusters in functional brain images. Neuroimage 8, 113–128 (1998)
6. Worsley, K., Marrett, S., Neelin, P., Vandal, A., Friston, K., Evans, A.C.: A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp 4, 58–73 (1996)
7. Holmes, A., Blair, R., Watson, J., Ford, I.: Nonparametric analysis of statistic images from functional mapping experiments. J. Cereb. Blood Flow Metab. 16(1), 7–22 (1996)
8. Nichols, T., Holmes, A.: Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain. Mapp. 15(1), 1–25 (2002)
9. Smith, S.M., Nichols, T.E.: Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44(1), 83–98 (2009)
10. Smith, S., Douaud, G., Salimi-Khorshidi, G., Webster, M., Mackay, C., Groves, A., Nichols, T.: Threshold-free cluster enhancement: Practical examples. In: 14th Anual Meeting of the Organization for Human Brain Mapping, Melbourne, Australia (2008)
11. Kiebel, S., Poline, J., Friston, K., Holmes, A., Worsley, K.: Robust smoothness estimation in statistical parametric maps using standardized residuals from the general linear model. Neuroimage 10(6), 756–766 (1999)
12. Flitney, D., Jenkinson, M.: Cluster analysis revisited. Technical report, FMRIB Centre, University of Oxford (2000)
13. Nichols, T.: Cluster analysis revisited - again: Implementing nonstationary cluster size inference. Technical report, FMRIB Centre, University of Oxford (2008)

Using Real-Time fMRI to Control a Dynamical System by Brain Activity Classification

Anders Eklund[1,2], Henrik Ohlsson[3], Mats Andersson[1,2], Joakim Rydell[1,2], Anders Ynnerman[4,2], and Hans Knutsson[1,2]

[1] Div. of Medical Informatics, Linköping University, Sweden
{andek,matsa,joary,knutte}@imt.liu.se
[2] Center for Medical Image Science and Visualization (CMIV)
[3] Div. of Automatic Control, Linköping University, Sweden
ohlsson@isy.liu.se
[4] Div. for Visual Information Technology and Applications
Linköping University, Sweden
andyn@itn.liu.se

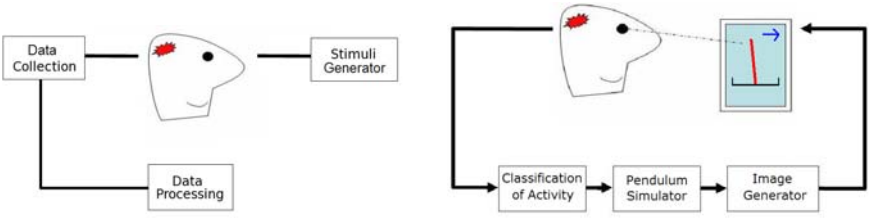
Abstract. We present a method for controlling a dynamical system using real-time fMRI. The objective for the subject in the MR scanner is to balance an inverted pendulum by activating the left or right hand or resting. The brain activity is classified each second by a neural network and the classification is sent to a pendulum simulator to change the force applied to the pendulum. The state of the inverted pendulum is shown to the subject in a pair of VR goggles. The subject was able to balance the inverted pendulum during several minutes, both with real activity and imagined activity. In each classification 9000 brain voxels were used and the response time for the system to detect a change of activity was on average 2-4 seconds. The developments here have a potential to aid people with communication disabilities, such as locked in people. Another future potential application can be to serve as a tool for stroke and Parkinson patients to be able to train the damaged brain area and get real-time feedback for more efficient training.

1 Introduction

Despite the enormous complexity of the human mind, fMRI, *functional Magnetic Resonance Imaging*, techniques are able to partially observe the state of a brain in action. In conventional fMRI, the experiment is performed and the data is analyzed afterwards to calculate a level of brain activity for each voxel. In real-time fMRI, the data is analyzed directly and the result from the analysis can be used to change the stimulus presented to the subject, see figure 1. The brain state can be interpreted by a computer and the setup is then often called a brain computer interface (BCI).

Brain computer interfaces first used electroencephalography (EEG). Some consider EEG to be superior to fMRI since it has a much higher temporal resolution. fMRI, however, has a higher spatial resolution than EEG. Even if the

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 1000–1008, 2009.



(a) A conventional fMRI experiment setup. The data is collected during the stimulus paradigm and is analyzed after the experiment to get a measure of brain activity in each voxel. The analysis can be made in real-time but is not used to change the stimulus given to the subject.

(b) The experiment setup for a bio-feedback loop using real-time fMRI. The subject sees the red pendulum. The activity for each new volume of data is classified and sent to the pendulum simulator. The blue arrow shows the current classification of the brain activity.

Fig. 1. The experiment setups for conventional fMRI and bio-feedback using real-time fMRI. In real-time fMRI we can achieve a bio-feedback loop, by using the result from the real-time analysis to change the stimulus, where the subject can control a dynamical system.

EEG signal is sampled with 5 kHz, most of the brain activity is below 100 Hz and an inverse problem has to be solved in order to reconstruct the spatial position of the brain activity. In most clinical applications, 19 recording electrodes are used according to the international 10-20 system [1]. The usable bandwidth of a normal clinical EEG application is thus 19 channels at 100 Hz, resulting in 1900 sample points per second. In our fMRI setup, we have 9000 channels (brain voxels) at 1 Hz, resulting in 9000 sample points per second. Faster scanning schemes using compressed sensing [2], can in the future increase the temporal resolution of fMRI. With stronger MR scanners it is also becoming easier to detect the small initial dip of the BOLD signal [3], and thus making it unnecessary to wait for the peak of the delayed BOLD response.

We have tested our BCI with a dynamical pole balancing experiment. In the experiment, the subject was given the possibility to move an inverted pendulum. The pendulum could be pushed to the left or right by activating the parts of the motor cortex associated with activity of the left and right hand. Similar projects have been presented in. [4] and [5]. Laconte et al. uses a similar setup in [5] but the arrow that they control can not be considered to be a dynamical system that changes by itself. The dynamical properties of the inverted pendulum make our setup a more challenging problem. We have to interpret the desire of the subject and set out a control signal moving the pendulum as often as once a second to have a chance to handle the fast dynamics of the pendulum.

The area of real-time fMRI is relatively new compared to conventional fMRI. Laconte et al. [5] were one of the first to perform real-time classification of brain activity. Ohlsson et al. [6] use real-time fMRI to determine where you are looking. deCharms et al. [7] use real-time fMRI for learning a subject to suppress its own pain. Weiskopf et al. [8] use real-time fMRI for self-regulation of local brain activity. An overview of fMRI brain computer interfaces is given by Sitaram et al. [9].

2 Methods

2.1 Measuring Brain Activity by Correlation

In fMRI the objective is to find the intensity and spatial position of brain activity. The estimation of brain activity is based on the fact that the magnetic properties of the blood changes when the neurons demand more oxygen to compensate for their increased activity. The body overcompensates the amount of oxygen sent to the neurons. The signal that is measured in fMRI is called the BOLD signal, where BOLD stands for *blood oxygen level dependent*.

In conventional fMRI, brain activity is often measured using correlation between the stimulus paradigm and the intensity time series of each voxel. A stimulus paradigm can for example consist of alternating between finger tapping and rest in periods of 20 seconds each. Since the subject is told what to do we calculate the brain activity for each voxel as the correlation with the stimulus paradigm. In order to improve the novel correlation method, several methods have been proposed. The most frequently used is the general linear model (GLM), e.g. Friston et al. [10].

2.2 Classification of Brain Activity

In our real-time fMRI setup, the subject acts independently of any paradigms. Thereby we do not have anything to correlate with. The BOLD signal that is detected in fMRI does not occur directly when some brain activity is started but is delayed 3-5 seconds. This is a property of the human physiology that we can not change. This makes it hard to control a system in real-time. In conventional fMRI it does not matter other than that the delay is unknown. Other difficulties in real-time fMRI are that it is harder, compared to conventional fMRI, to detrend the time series of each voxel and that all calculations have to be made in real-time. Detrending is needed since there are drifts and trends in the fMRI data that will corrupt the estimates if not removed [11].

Instead of using correlation, we classify each volume of data to know what the current brain activity is. In this project we classify between left hand activity, right hand activity and rest. A training phase is used to learn how to classify between the different types of brain activity. We used a one layer neural network to classify the brain activity as a first approach since it is easy to implement and to explain and provided sufficient discrimination

2.3 Neural Networks

Neural networks [12] are used in many applications to classify data into a number of predetermined classes. Often some informative properties are first extracted from the data instead of using the raw data itself in the classification. Our first approach was to find active voxels in the left and right motor cortex and use them for the classification, to reduce the dimension of the original data. However, it is a known fact that even if the subject is only activating one hand, there will

be activity in both the left and right part of the motor cortex. Hence it is hard to create a stimulus paradigm for the training phase of the classifier to find voxels that only are active in each state. In the more simple case with only two states, rest and activity, we can find the best voxels, by correlation, to use in the classification.

Neural networks can have arbitrary many layers of perceptrons. A one layer neural network is sufficient if the classes are linearly separable, otherwise at least two layers have to be used in order to classify the data. The best weights w of the discriminant function is derived during the training phase where training vectors and the according correct classifications are presented to the neural network. In our training phase the subject follows a pre-determined stimulus paradigm. We trained the neural network with all the brain voxels in each collected volume, using a spatial mask. Normally 9000 voxels, out of 64000, were considered to be brain voxels. Each volume was also filtered by a 3 x 3 x 3 Gaussian lowpass filter to make the classifications more robust to motions.

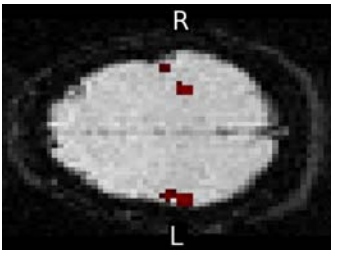
We used a one layer neural network with 3 output nodes. Each output node has a weight vector of dimension 9000 that is achieved during the training phase. In the training phase we used a 240 second long stimulus paradigm consisting of 20 seconds of left activity, followed by 20 seconds of rest and then 20 seconds of right activity, repeated 4 times. A volume of data was collected each second resulting in 240 training vectors, each of dimension 9000, for the neural network.

Since the dimension of the data is 9000 and the number of training vectors is only 240, we use a linear activation function during the training of the neural network. This means that our neural network is equivalent to using linear regression for classification. If for example *tanh* would have been used as an activation function instead, the risk of overfitting would have been large since the number of dimensions is much higher than the number of training examples. To make a classification, three scalar products are calculated and the class corresponding to the output node with the highest output, meaning the furthest distance from the decision boundary, is selected. In the real-time phase, each volume was spatially smoothed (3 x 3 x 3 Gaussian lowpass filter) and detrended (by removing the mean value calculated 45 seconds backwards for each voxel separately). A control signal (apply a force to the left, right or do nothing) was then computed by applying the, in the training phase computed, neural network. The control signal was used as an input to the simulation of the dynamical pendulum system. The simulation uses the control signal to change the state of the system by applying an according force to the pendulum.

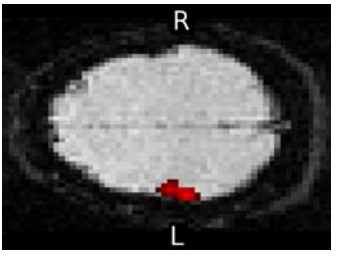
2.4 Head Movement

One potential cause of error in fMRI is that the subject may involuntarily move the head in pace with the stimulus paradigm and thereby induce high correlation in voxels on the edge of the brain or outside the brain.

For a neural network it is straight forward to calculate the importance of each voxel for each classification. From this we can eliminate the possibility that the reason that our experiment setup works is due to head movement.



(a) The most important voxels for classification of left hand activity at one timepoint.



(b) The most important voxels for classification of right hand activity at one timepoint.

Fig. 2. The figures show which voxels that are the most important for classification of brain activity at two timepoints. The importance maps should not be interpreted as activity maps. It is clear that the voxels in the motor cortex, as expected, are the most important. If head movement was the reason that the experiment setup worked, the voxels at the edge of the brain would have been important for each classification. The explanation for the overlap for left and right hand activity classification, which is not present for all timepoints, is that the information that there is no activity in one side of the motor cortex is important for the classification and that bilateral activation of the motor cortex is common (but different for left and right hand activation). We use the radiological display convention and have marked the left and the right side with L and R.

If the distance to the decision boundary, for a classification at one timepoint t, is denoted with d_t, then the importance i of voxel v_{nt} (voxel v_n at timepoint t) can be calcuclated as

$$i_{v_{nt}} = \frac{w_n v_{nt}}{d_t} = \frac{w_n v_{nt}}{\sum_k w_k v_{kt}} \tag{1}$$

where w_n is the weight in the neural network for voxel v_n. Before the calculation is made, one has to make sure that the voxel contributes to the right direction from the decision boundary, i.e. that $w_n v_{nt} > 0$. If these calculations are made for all the brain voxels, there will be a lot of voxels that are important for some timepoint but not for a set of continuous timepoints. In order to remove the flickering voxels, a median filtering of the time series of the importance value for each voxel can be made. The resulting importance maps are shown in figure 2. A 3 x 3 gaussian lowpass filter has been used to improve the appearance of the importance maps.

3 Experiment Setup

The data was acquired using a 1.5 T Philips Achieva MR scanner. The acquisition resolution was 80 x 80 x 10 voxels. Field of view and slice thickness were chosen to obtain a voxel size of 3 x 3 x 3 mm. Echo time (TE) was set to 40 ms and repetition time (TR) was set to 1000 ms. The classification of the brain activity and the simulation of the inverted pendulum was carried out in Matlab on a

standard laptop. The current state of the inverted pendulum was shown to the subject in a pair of VR goggles.

The resulting angular acceleration was calculated in each iteration and we used a small timestep to calculate the angular velocity and the angle of the pendulum. The simulation of the pendulum was carried out continously and each time we collected a new volume of data (once a second), a classification of brain activity was made and used to set the force applied to the pendulum.

4 Results

The subject was able to balance the inverse pendulum both with real activity and imagined activity. When the subject used imagined activity, the pendulum was balanced by only thinking of activating the left or right hand. To justify the success of the controller figure 3 shows the logged angle of the pendulum and the classified activity during the real-time phase. The figure shows that the subject could balance the pendulum for 250 s. The figure also shows the dynamical properties of the pendulum. To validate that this significantly differs from chance, 5000 simulation runs of 800 s each was performed with the pendulum simulator with random activity. 35 out of 59446 simulated balancing attempts (0.059%) were then longer than 250 s.

We also tested the neural network offline with four different datasets, two generated by real activity (R1, R2) and two generated by imagined activity (I1, I2). All datasets were 240 s long. The detrending was performed as in the real-time phase. The number of correct classifications, as function of training dataset and evaluation dataset used, is given in table 1 below. A phase based 3D rigid body motion compensation algorithm was implemented. It did not run in real-time during the experiment but was used offline for validation.

Table 1. Neural network classification performance as function of training dataset and evaluation dataset

Training dataset / Evaluation dataset	R1/R2	R2/R1	I1/I2	I2/I1	R1/I1	R1/I2
Without motion compensation	93.7%	86.3%	86.7%	85.8%	90.4%	83.3%
With motion compensation	93.4%	90.4%	86.3%	86.7%	90.8%	83.8%
Training dataset / Evaluation dataset	R2/I1	R2/I2	I1/R1	I1/R2	I2/R1	I2/R2
Without motion compensation	87.5%	77.1%	90.0%	90.0%	89.6%	91.7%
With motion compensation	87.1%	77.9%	89.2%	88.8%	87.1%	91.7%

It is worth to mention that classification with ordinary linear regression performs as well as the far more advanced support vector machines (SVM) approach used by Laconte et al. [5].

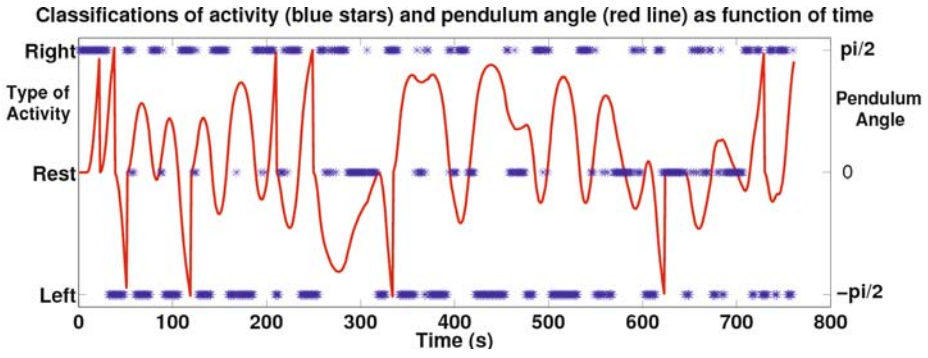


Fig. 3. The figure shows the angle of the pendulum, the red line, and the classifications of activity, the blue stars, as a function of time in the real-time phase. If the angle exceeds $-\frac{\pi}{2}$ or $\frac{\pi}{2}$ the pendulum is restarted and the angle is set to zero, this happened 9 times in this real-time phase. The test subject has to compensate for the delayed BOLD-signal by changing activity a few seconds in advance. It is easy to see that the test subject needed a couple of attempts first to learn the dynamics of the system. The effective force applied to the pendulum depends on the angle of the pendulum as $cos(\alpha)$. Angle 0 means that the pendulum is standing straight up, $-\frac{\pi}{2}$ means that the pendulum lies along the negative x-axis and $\frac{\pi}{2}$ means that the pendulum lies along the positive x-axis. This means that it is easy to move the pendulum when it is standing straight up, but it takes a long time to straighten it up if the angle is close to $-\frac{\pi}{2}$ or $\frac{\pi}{2}$. If the activity is classified as left activity, a force to the left on the pendulum is applied and the angle of the pendulum decreases. If the activity is classified as right activity, a force to the right is applied and the angle of the pendulum increases. If the activity is classified as rest, no force is applied (other than the gravitational force) and the pendulum continues to rotate in its current angular direction. Between approximately timepoint 350 and 600 in the figure the pendulum was successfully balanced by the subject.

5 Discussion

We have presented an fMRI based BCI realization. The human brain and a computer were here linked by fMRI and worked together as a controller of a dynamical system. The dynamical system was made up of an inverted pendulum. The subject had the ability to induce a force by evoking brain activity in the motor cortex. A neural network was trained to separate between left and right hand activity and rest. The subject was able to balance the inverted pendulum both with real activity and with imagined activity. In the future we would like to improve the detection speed of the system. One way to do this is to train the classifier on the transitions between the different states instead of the states them self, as mentioned in [5]. This can be done by looking back at the signal a number of seconds to learn what the different transitions look like, to earlier detect a change of activity. We would also like to increase the bandwidth of the bio-feedback loop by including a larger number of different simultaneous activities

to make the subject able to control a more advanced dynamical system. We also want to make our motion compensation algorithm to run in real-time and compare it to existing techniques [13].

The developments here can be used for neuro-scientific investigations [14] and also have a potential to aid people with communication disabilities, such as locked in people [15]. Other possible applications are to learn how control your own pain [7] or to serve as a tool for stroke [16] and Parkinson [17] patients to be able to train the damaged brain area and get real-time feedback for more efficient training.

Acknowledgments. This work was supported by the Strategic Research Center MOVIII, funded by the Swedish Foundation for Strategic Research, SSF.

References

1. Niedermeyer, E., da Silva, L.: Electroencephalography: Basic Principles, Clinical Applications, and Related Fields, 5th edn. Lippincott Williams & Wilkins (2004) ISBN 0781751268
2. Lustig, M., Donoho, D., Pauly, J.M.: Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine 58, 1182–1195 (2007)
3. Yacoub, E., Hu, X.: Detection of early negative response in fMRI at 1.5 tesla. Magnetic Resonance in Medicine 41, 1088–1092 (1999)
4. Yoo, S., Fairneny, T., Chen, N., Choo, S., Panych, L.P., Park, H., Lee, S., Jolesz, F.A.: Brain-computer interface using fMRI: spatial navigation by thoughts. Neuroreport 15, 1591–1595 (2004)
5. Laconte, S.M., Peltier, S.J., Hu, X.P.: Real-Time fMRI Using Brain-State Classification. Human Brain Mapping 28, 1033–1044 (2007)
6. Ohlsson, H., Rydell, J., Brun, A., Roll, J., Andersson, M., Ynnerman, A., Knutsson, H.: Enabling Bio-Feedback Using Real-Time fMRI. In: 47th IEEE Conference on Decision and Control (2008)
7. de Charms, R.C., Maeda, F., Glover, G.H., Ludlow, D., Pauly, J.M., Soneji, D., Gabrieli, J.D.E., Mackey, S.C.: Control over brain activation and pain learned by using real-time functional MRI. PNAS 102, 18626–18631 (2005)
8. Weiskopf, N., Scharnowski, F., Veit, R., Goebel, R., Birbaumer, N., Mathiak, K.: Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). Journal of Physiology-Paris 98, 357–373 (2004)
9. Sitaram, R., Caria, A., Veit, R., Gaber, T., Rota, G., Kuebler, A., Birbaumer, N.: fMRI Brain-Computer Interface: A Tool for Neuroscientific Research and Treatment. Computational Intelligence and Neuroscience (2007)
10. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C., Frackowiak, R.S.J.: Statistical Parametric Maps In Functional Imaging: A General Linear Approach. Human Brain Mapping 2, 189–210 (1995)
11. Friman, O., Borga, M., Lundberg, P., Knutsson, H.: Detection and detrending in fMRI data analysis. NeuroImage 22, 645–655 (2004)
12. Haykin, S.: Neural networks - a comprehensive foundation. Prentice-Hall, Englewood Cliffs (1999)

13. Oakes, T.R., Johnstone, T., Walsh, K.S.O., Greischar, L.L., Alexander, A.L., Fox, A.S., Davidson, R.J.: Comparison of fMRI motion correction software tools. NeuroImage 28, 529–543 (2005)
14. Hollmann, M., Baecke, S., Mueller, C., Bernarding, J.: Predicting Human Decisions in a Social Interaction-Scenario Using Real-Time Functional Magnetic Resonance Imaging (Rt-FMRI), ISMRM (2009)
15. Boly, M., Coleman, M.R., Davis, M.H., Hampshire, A., Bor, D., Moonen, G., Maquet, P.A., Pickard, J.D., Laureys, S., Owen, A.M.: When thoughts become action: An fMRI paradigm to study volitional brain activity in non-communicative brain injured patients. NeuroImage 36, 979–992 (2007)
16. Carey, J.R., Kimberley, T.J., Lewis, S.M., Auerbach, E.J., Dorsey, L., Rundquist, P., Ugurbil, K.: Analysis of fMRI and finger tracking training in subjects with chronic stroke. Brain 125, 773–788 (2002)
17. Wu, T., Hallett, M.: A functional MRI study of automatic movements in patients with Parkinson's disease. Brain 128, 2250–2259 (2005)

Modeling Adaptation Effects in fMRI Analysis

Wanmei Ou[1], Tommi Raij[2], Fa-Hsuan Lin[2,3], Polina Golland[1], and Matti Hämäläinen[2]

[1] Computer Science and Artificial Intelligence Laboratory, MIT, USA
[2] Athinoula A. Martinos Center for Biomedical Imaging, MGH, USA
[3] Institute of Biomedical Engineering, National Taiwan University, Taiwan

Abstract. The standard general linear model (GLM) for rapid event-related fMRI design protocols typically ignores reduction in hemodynamic responses in successive stimuli in a train due to incomplete recovery from the preceding stimuli. To capture this adaptation effect, we incorporate a region-specific adaptation model into GLM. The model quantifies the rate of adaptation across brain regions, which is of interest in neuroscience. Empirical evaluation of the proposed model demonstrates its potential to improve detection sensitivity. In the fMRI experiments using visual and auditory stimuli, we observed that the adaptation effect is significantly stronger in the visual area than in the auditory area, suggesting that we must account for this effect to avoid bias in fMRI detection.

1 Introduction

Rapid event-related (ER) functional magnetic resonance imaging (fMRI) is one of the most popular imaging methods in cognitive neuroscience. In the rapid ER fMRI studies, individual stimuli are presented every few seconds or faster. Although less efficient for localizing activation, rapid ER fMRI has several advantages over the traditional block design, including the ability to randomize trial types and to sort data based on behavioral responses.

The standard analysis for rapid ER fMRI models activation as a linear system [2,5,9]; the hemodynamic response to multiple input stimuli is assumed to be a superposition of the responses to individual stimuli. This approach estimates the impulse response function, also known as the hemodynamic response function (HRF), of this linear system via de-convolution, and compares the estimates to the null hypothesis or to estimates from other experimental conditions.

fMRI signals commonly do not comply with the linear assumption. Independent studies have demonstrated a substantial adaptation effect in the hemodynamic response [1,11,12,16,17], i.e., if two stimuli are presented within the adaptation period, the amplitude of the response to the second stimulus is reduced. Furthermore, the adaptation effect strengthens as the inter-stimulus interval (ISI) decreases. Several studies demonstrated that when a pair of visual stimuli is presented less than 1 sec apart, the response amplitude to the second

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 1009–1017, 2009.

stimulus is approximately 55% of that to the first stimulus, with recovery to 90% at a 6 sec ISI [12,16,17]. This evidence suggests that the adaptation effect must be modeled in the analysis, especially when stimuli are presented frequently.

The adaptation effect is expected to vary spatially due to differences in neural and hemodynamic properties of functional areas in the brain [1,13,16]. While physiological mechanisms for adaptation are not clearly understood, it is still useful to model it for the purposes of improving detection.

Previous studies of the adaptation effect separated detection and adaptation modeling [12,13,15,16,17], fitting the adaptation model to the estimated HRF obtained using the standard general linear model (GLM) [9]. This approach ignores the trial-to-trial variation. Work by Buxton *et al.* [4] introduced the biophysical balloon model for fMRI signals where the adaptation effect is captured through interactions among blood flow, blood volumes, and de-oxyhemoglobin content, instead of an explicit interaction between stimuli. Friston *et al.* [10] proposed a statistical model using the Volterra kernels to capture interaction between stimuli. The interaction can be efficiently estimated and statistically examined via the F-test. However, the physiological interpretation of the model parameters is challenging, since the model treats the stimuli symmetrically, effectively ignoring the causal nature of the adaptation effect.

In this work, we extend the basic GLM by incorporating a region-specific model of adaptation. In addition to the stimulus onset, our design matrix also depends on the ISIs between stimuli via a single-parameter exponential function. Specifically, this model captures the decrease in the magnitude of the hemodynamic response if the time interval to the preceding stimuli is short. In other words, we only model causal interactions among stimuli, in contrast to the bi-directional interaction model in [10]. By combining detection and adaptation modeling, the proposed method takes into account trial-to-trial variation. It is expected that the adaptation effect strengthens when more stimuli are presented prior to the current stimulus. We summarize this effect from multiple stimuli through a multiplicative model. This extension allows for a more flexible choice of an experimental paradigm in contrast to previous fMRI adaptation studies which were restricted to presentations of stimulus pairs [12,13,16,17].

We jointly estimate the decay parameter of the exponential function for each region and the HRF for each location in the brain. The estimated parameter of the exponential function reflects the length of the adaptation period for the corresponding region, and the estimated HRF indicates the activation status of the corresponding location. Our experimental results demonstrate a significant improvement in detection sensitivity and confirm previously known adaptation phenomena in the sensory systems.

2 Method

The univariate GLM [9] assumes that the fMRI signal $\boldsymbol{y}_n$ at location n is the superposition of the hemodynamic responses to the stimuli in the paradigm, of physiological signal, and of noise:

$$\boldsymbol{y}_n = \mathbf{X}\boldsymbol{\beta}_n + \boldsymbol{\Phi}\boldsymbol{\alpha}_n + \boldsymbol{\epsilon}_n, \tag{1}$$

where $\mathbf{X}$ is the design matrix, constructed based on the experimental protocol. Columns of matrix $\boldsymbol{\Phi} = [\phi_1, \cdots, \phi_R]$, often in the form of low-order polynomials, model the protocol-independent factors such as cardiac activity and breathing. The measurement noise $\boldsymbol{\epsilon}_n$ can be modeled as white Gaussian noise or as colored noise with an auto-regressive (AR) structure [3]. Vectors $\boldsymbol{\beta}_n$ and $\boldsymbol{\alpha}_n$ are the corresponding coefficients of the protocol-dependent and the protocol-independent signals, respectively. To determine the activation status at location n, one compares the estimated protocol-dependent coefficient $\widehat{\boldsymbol{\beta}}_n$ to the null hypothesis or to the corresponding estimates for other experimental conditions.

Without loss of generality, we assume a single type of stimulus. The matrix form of GLM represents the convolution of experimental protocol and HRF:

$$y_n(t) = \sum_{k=1}^{K} h(t - s_k) \star \beta_n(t) + \sum_{r=1}^{R} \alpha_{rn}\phi_r(t) + \epsilon_n(t), \tag{2}$$

where $\boldsymbol{s} = [s_1, s_2, \cdots, s_K]$ is the vector of onset times of the K stimuli in the experiment. When modeling the HRF using a finite impulse response (FIR) model, h is the impulse train, and β_n contains the values of the FIR model. When modeling HRF with a fixed kernel, h is the convolution of the stimulus train and the kernel, and β_n is a scalar that modulates the HRF magnitude.

The above model fails to capture the fact that previous stimuli can decrease the hemodynamic response to subsequent stimuli if the recovery period is longer than the ISI presented. Therefore, we incorporate an adaptation model into the standard GLM by introducing a damping weight w_k for each stimulus. Due to the regionally varying neuronal and vascular architecture of the brain, we parameterize the weight $w_k(\boldsymbol{s}; \theta_m)$ with a region-specific parameter, i.e., θ_m for region m. $w_k(\boldsymbol{s}; \theta_m)$ accumulates the adaptation effect from stimuli prior to stimulus k, presented at time s_k. Therefore, in the new model, the fMRI signal at location n of region m is the superposition of the weighted version of the response to each stimulus:

$$y_n(t) = \sum_{k=1}^{K} [w_k(\boldsymbol{s}; \theta_m) h(t - s_k)] \star \beta_n(t) + \sum_{r=1}^{R} \alpha_{rn}\phi_r(t) + \epsilon_n(t). \tag{3}$$

In the matrix notation this equation reads

$$\boldsymbol{y}_n = \widetilde{\mathbf{X}}(\theta_m)\boldsymbol{\beta}_n + \boldsymbol{\Phi}\boldsymbol{\alpha}_n + \boldsymbol{\epsilon}_n \quad \forall\ n \in V_m, \tag{4}$$

where V_m is the set of locations in region m. The new design matrix $\widetilde{\mathbf{X}}$ depends not only on the stimulus onset times, but also on the ISIs between stimuli. Compared to the standard GLM model in Eq. (1), the nonlinear effect is captured in the design matrix $\widetilde{\mathbf{X}}$.

Adaptation model. We combine the adaptation effects for stimulus k using a multiplicative exponential model, ranging between zero and one:

$$w_k(\boldsymbol{s}; \theta_m) = \prod_{j=1}^{k-1} \left(1 - e^{-\theta_m(s_k - s_j)}\right). \tag{5}$$

The exponential decay parameter θ_m models the length of the adaptation period. A larger value of θ_m indicates a weaker adaptation effect, or a shorter period for

the region to recover. The multiplicative nature of the model reflects the fact that multiple preceding stimuli can affect the amplitude of the response to a particular stimulus. In other words, the brain response is modeled as a Markov process of order K.

Inference. We estimate the unknown parameters $\{\theta_m, \{\boldsymbol{\beta}_n, \boldsymbol{\alpha}_n\}_{n\in V_m}\}$ in Eq. (4) by minimizing the sum of squares of the residual errors for each region independently. For a given θ_m, the optimal linear parameters $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\alpha}}_n$ can be found in a closed-form:

$$\left[\widehat{\boldsymbol{\beta}}_n, \widehat{\boldsymbol{\alpha}}_n\right] = \left(\mathbf{H}^{\mathrm{T}}(\theta_m)\mathbf{H}(\theta_m)\right)^{-1}\mathbf{H}^{\mathrm{T}}(\theta_m)\boldsymbol{y}_n, \tag{6}$$

where $\mathbf{H}(\theta_m) = [\widetilde{\mathbf{X}}(\theta_m)\ \boldsymbol{\Phi}]$. Substituting Eq. (6) into the expression for the residual error, we obtain the optimal $\widehat{\theta}^*_m$:

$$\widehat{\theta}^*_m = \arg\min_{\theta} \textstyle\sum_{n\in V_m} \left\| \left(\mathbf{I} - \left(\mathbf{H}^{\mathrm{T}}(\theta)\mathbf{H}(\theta)\right)^{-1}\mathbf{H}^{\mathrm{T}}(\theta)\right)\boldsymbol{y}_n \right\|^2. \tag{7}$$

With the proposed adaptation model, Eq. (7) is a nonlinear function of a single scalar parameter θ_m. We can simply exhaustively search for the parameter value within a specified range. We then obtain the optimal values $\widehat{\boldsymbol{\beta}}^*_n$ and $\widehat{\boldsymbol{\alpha}}^*_n$ by substituting $\widehat{\theta}^*_m$ into Eq. (6).

Due to the nonlinearity of the model, the true value of θ is needed to compute the covariance of the estimate, $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}^*_n)$. Since θ is not known in real experiments, we approximate $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}^*_n)$ with $\mathrm{Cov}(\widehat{\boldsymbol{\beta}}^*_n; \widehat{\theta}^*_m)$. Hence, the statistic $\widehat{\boldsymbol{\beta}}^{*\mathrm{T}}_n \mathrm{Cov}^{-1}(\widehat{\boldsymbol{\beta}}^*_n; \widehat{\theta}^*_m)\widehat{\boldsymbol{\beta}}^*_n / N_{\boldsymbol{\beta}_n}$, where $N_{\boldsymbol{\beta}_n}$ is the number of regression coefficients in $\boldsymbol{\beta}_n$, does not follow a known probability distribution under the null hypothesis, in contrast to the F-distribution in the standard GLM analysis. This represents a challenge in testing significance similar to GLM with AR noise modeling [3]. The exact statistical test can be achieved with Markov Chain Monte Carlo simulation, but it is too computationally intensive for a standard analysis procedure. We will see in the next section that comparing the values of the statistic across locations provides insight into the adaptation effect. Developing efficient methods for assessing statistical significance of the statistic is clearly a direction for future research.

To summarize, by accounting for the adaptation effects, we obtain a more accurate estimate of the HRF which leads to more accurate detection results. The estimates of the adaptation parameter θ promise to provide an insight into the neuronal and vascular architecture across different brain regions. In the following, we refer to our approach as GLMA (GLM with adaptation).

3 Results

Due to the lack of ground truth in real experiments, we first study GLMA's sensitivity to noise and parameter settings using simulated data. We then compare GLMA to GLM using human fMRI data from a visual-auditory study.

In both simulations and analysis of human fMRI data, we constrain the detection to the cortex and define different brain regions based on the cortical folding

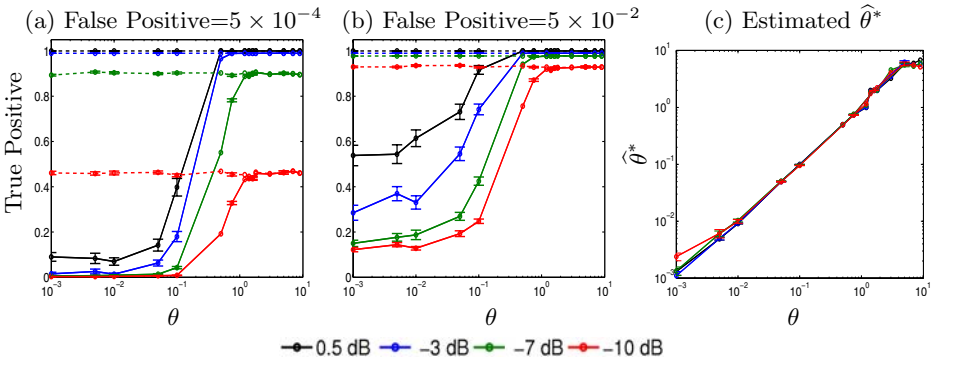


Fig. 1. The two panels on the left compare the true positive rates for GLM (solid) and GLMA (dashed) as a function of θ, for two false positive rates and four SNR values. The rightmost panel shows the estimates $\widehat{\theta}^*$ of the adaptation parameter θ. The standard error bars of the true positive rates and estimates $\widehat{\theta}^*$ are presented in the corresponding panels.

patterns, obtained with the FreeSurfer software [6,8], 35 regions per hemisphere. Moreover, since the adaptation weight w_k recovers exponentially with respect to ISI, we only consider the stimulus interactions within a 16 sec window.

Simulation studies. Since our model is estimated for each region separately, it is sufficient to study the performance of the model for a range of values of θ using data in a single region. For each value of θ, we generated 100 activation time courses by convolving a two-gamma function [14] with a train of stimuli whose onset times were generated randomly (mean ISI=4.0 sec, std=3 sec). The adaptation effect was modeled according to Eq. (5). We also generated 100 time courses without activation. We then added i.i.d. Gaussian noise to these 200 time courses, and repeated the simulation procedure 50 times for each value of θ. We chose noise levels to be within the typical SNR range of real fMRI data.

We separately processed the data set using GLM and GLMA. In both cases, HRF was modeled with the two-gamma function used in the simulation. Thus, β is a scalar in this case. Based on the activation statistic $(\widehat{\beta}^*)^2/\mathrm{Var}\left(\widehat{\beta}^*\right)$, we evaluated the detection accuracy of both methods at two false positive rates, 5×10^{-4} and 5×10^{-2}, as shown in Fig. 1(a,b). As expected, a better SNR in the data allows for a higher detection rate over the range of θ we examined. When θ is larger than 0.5, there is essentially no adaptation effect present in the data. Hence, the performance of the two methods is almost identical. The adaptation effect strengthens as θ decreases. We can clearly see an up-to 80% higher detection rate achieved by modeling the effect.

We also investigated the robustness of the estimation of the adaptation parameter θ. As illustrated in Fig. 1(c), the estimates $\widehat{\theta}^*$ closely match the simulation. As θ decreases, the response to the subsequent stimuli is very small for the chosen mean ISI, and $\widehat{\theta}^*$ starts to deviate from the true value θ for noisy data, i.e., SNR$= -10$ dB.

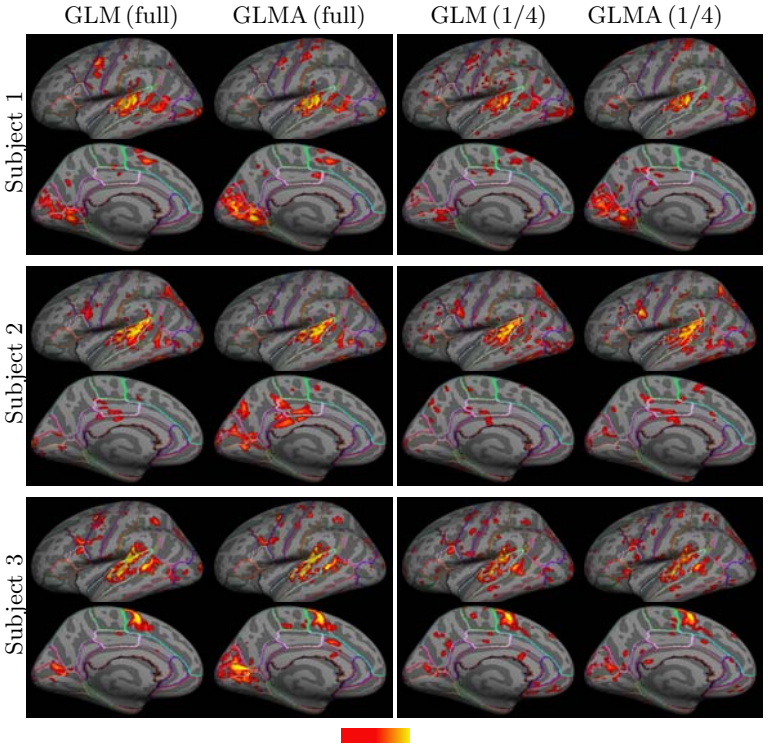


Fig. 2. The thresholded statistical parametric maps for three subjects using GLM with full length data (first row) and with 1/4 of the data (third row), as well as using GLMA with full-length data (second row) and with 1/4 of the data (fourth row)

Human experiments. We illustrate the application of the proposed method in a visual-auditory fMRI study. In this experiment, three healthy right-handed subjects were presented with visual-auditory stimuli in a random rapid ER fMRI paradigm in three runs, with mean ISI of 1.5 sec (std=0.7 sec), 3.0 sec (std=1.3 sec), and 6.0 sec (std=2.1 sec), respectively. To keep subjects engaged throughout the experiment, they were asked to respond to a rare target stimulus by pressing a button. The fMRI data were acquired using a 3T Siemens Trio scanner (TR 1.15 sec, TE=30 msec, flip angle 90 degree, $3.1 \times 3.1 \times 4\ \text{mm}^3$). Structural T1-weighted MRI scans of the subjects were acquired with a 1.5T Siemens Avanto scanner. The anatomical images were processed with the FreeSurfer software [6,8]. Individual subject functional scans were morphed through a spherical mapping into an atlas constructed with 40 subjects [7].

We applied GLM and GLMA to data combined from all three runs (Fig. 2 left) and data combined from the first quarter of each of the three runs (Fig. 2 right). Since the statistics across methods are not directly comparable, and the ground truth activation is not known, we select top 5% of vertices in a hemisphere with the highest statistics, and visually evaluate the results to assess the importance of modeling the adaptation effect. We emphasize that in order to

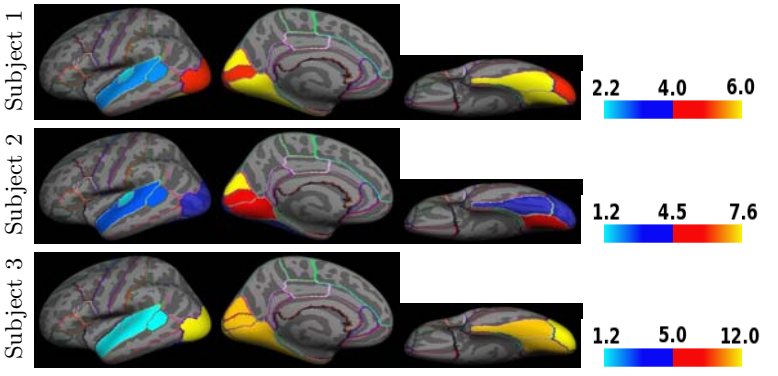


Fig. 3. Estimated 90% recovery period, t_{90}, for the three subjects presented in Fig. 2

develop valid detectors, further work is required in building statistical tests for assessing significance of the detected activations.

Fig. 2 shows that the two methods provide similar detection results in the auditory area. However, the visual region detected by GLM has a smaller spatial extent than the corresponding results using GLMA. Without modeling the adaptation effects, the activation statistics in the visual area are smaller than those in the auditory area. Therefore, many activations on the visual cortex will be missed if a single threshold is applied to the whole brain. Furthermore, compared to GLM, our detection results with shorter length data are more similar to the results using full-length data, indicating improved robustness.

For the selected auditory and visual areas, we present the time required for the attenuation coefficient w_k to recover to 90%, i.e., $t_{90} = -\theta^{-1} \ln(1 - 0.9)$, assuming a single stimulus presented prior to the current one (Fig. 3). Across all three subjects, t_{90} in the auditory areas is shorter than that in the visual areas, reflecting a stronger adaptation effect in the visual areas than in the auditory areas. Furthermore, in Subjects 1 and 2, the early visual regions, such as the lateral-occipital area, exhibit a weaker adaptation effect than high-order visual regions, such as the lingual and fusiform areas. For these two subjects, t_{90} is about 6 sec at the calcarine area which agrees with findings reported in [12,16,17]. The adaptation effects in Subject 3 are significantly longer than those in the other two subjects. Further experiments are needed to better understand and model the variability of the effect across subjects. To further validate our method, we also applied GLM, with HRF modeled as a two-gamma function, separately to each of the three runs. Fig. 4 shows the average estimated HRFs for the 50 most active vertices in each of the three selected regions for Subject 2. We can clearly see that the lingual area exhibits the strongest adaptation effect ($t_{90} =$ 5.8 sec in Fig. 3), indicated by substantial differences in response magnitude to stimuli presented with different mean ISI conditions. The difference in response magnitudes to the three conditions is smaller in the lateral-occipital area ($t_{90} =$ 3.8 sec), reflecting a weaker adaptation. On the other hand, the estimated HRFs across different ISIs are almost identical in the superior-temporal area ($t_{90} =$ 2.3 sec). That means the most active locations in the superior-temporal area

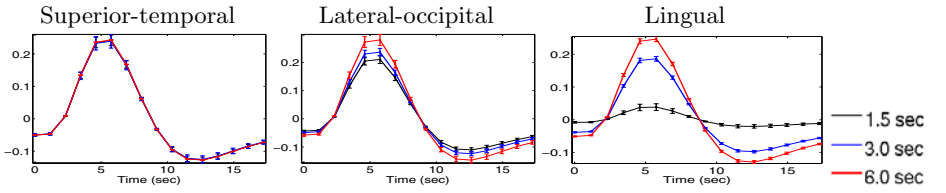


Fig. 4. Estimated HRFs for stimuli presented at three different mean ISI conditions, 1.5, 3.0, and 6.0 sec, obtained with standard GLM analysis in Subject 2. The error bars indicate standard deviation of the 50 selected vertices.

can almost fully recover in about 1.5 sec. The results of the HRF analysis for separate conditions agrees with our estimates of the adaptation effects in Fig.3.

4 Conclusions

We proposed and demonstrated a novel method for modeling the adaptation effects in fMRI. Experimental results indicate a significant improvement in detection sensitivity. The proposed method also quantifies the adaptation effects across brain regions, providing insight into neuronal and vascular organization of the brain. The current adaptation model focuses on the ISIs. We plan to extend it to include information about the amplitude of the response to previous stimuli, since the adaptation effect is expected to be more pronounced immediately after a strong response than after a weak response. Our framework can be readily modified to include different functional forms of the magnitude changes due to adaptation. We will explore alternative functions, such as the sigmoid function, in adaptation modeling in future work.

Acknowledgments. This work was supported in part by NIH R01 NS048279, NIH R01 EB006385, NIH NIBIB NAMIC U54-EB005149, NIH NCRR NAC P41-RR13218, NIH NCRR P41-RR14075 grants, NSF CAREER Award 0642971, Sigrid Juselius Foundation, Academy of Finland, and Finnish Cultural Foundation. Wanmei Ou is partially supported by the PHS training grant DA022759-03.

References

1. Birn, R.M., Saad, Z.S., Bandettini, P.A.: Spatial heterogeneity of the nonlinear dynamics in the fMRI BOLD response. NeuroImage 14, 817–826 (2001)
2. Boynton, G.M., Engel, S.A., Glover, G.H., Heeger, D.J.: Linear systems analysis of functional magnetic resonance imaging in human V1. J. NeuroSci. 16, 4207–4221 (1996)
3. Burock, M., Dale, A.: Estimation and detection of event-related fMRI signals with temporally correlated noise: a statistically efficient and unbiased approach. Hum. Brain Mapp. 11, 249–260 (2000)
4. Buxton, R.B., Wong, E.C., Frank, L.R.: Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn. Reson. Med. 39, 855–864 (1998)

5. Dale, A.M., Buckner, R.: Selective averaging of rapidly presented individual trials using fMRI. Hum. Brain Mapp. 5, 329–340 (1997)
6. Dale, A.M., Fischl, B., Sereno, M.I.: Cortical surface-based analysis: I. segmentation and surface reconstruction. NeuroImage 9, 179–194 (1999)
7. Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J.: An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31, 968–980 (2006)
8. Fischl, B., Sereno, M., Dale, A.M.: Cortical surface-based analysis: II. inflation, flattening, and a surface-based coordinate system. NeuroImage 9, 195–207 (1999)
9. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-B., Frith, C.D., Frackowiak, R.: Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Mapp. 2, 189–210 (1994)
10. Friston, K.J., Mechelli, A., Turner, R., Price, C.J.: Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12, 466–477 (2000)
11. Heckman, G.M., Bouvier, S.E., Carr, V.A., Harley, E.M., Cardinal, K.S., Engel, S.A.: Nonlinearities in rapid event-related fMRI explained by stimulus scaling. NeuroImage 34, 651–660 (2007)
12. Huettel, S., McCarthy, G.: Evidence for a refractory period in the hemodynamic response to visual stimuli as measured by MRI. NeuroImage 11, 547–553 (2000)
13. Huettel, S., McCarthy, G.: Regional differences in the refractory period of the hemodynamic response: an event-related fMRI study. NeuroImage 14, 967–976 (2001)
14. Jezzard, P., Matthews, P.M., Smith, S.M.: Functional MRI: an introduction to methods. Oxford University Press, Oxford (2002)
15. Robson, M.D., Dorosz, J.L., Gore, J.C.: Measurements of the temporal fMRI response of the human auditory cortex to trains of tones. NeuroImage 7, 185–198 (1998)
16. Soon, C.S., Venkatraman, V., Chee, M.W.: Stimulus repetition and hemodynamic response refractoriness in event-related fMRI. Hum. Brain Mapp. 20, 1–12 (2003)
17. Zhang, N., Zhu, X.H., Chen, W.: Investigating the source of BOLD nonlinearity in human visual cortex in response to paired visual stimuli. NeuroImage 43, 204–212 (2008)

A Cluster Overlap Measure for Comparison of Activations in fMRI Studies

Guillermo A. Cecchi, Rahul Garg, and A. Ravishankar Rao

IBM T. J. Watson Research Center, NY, USA

Abstract. Most fMRI studies use voxel-wise statistics to carry out intra-subject as well as inter-subject analysis. We show that statistics derived from voxel-wise comparisons are likely to be noisy and error prone, especially for inter-subject comparisons. In this paper we propose a novel metric called *weighted cluster coverage* to compare two activation maps. This metric is based on the intersection of spatially contiguous clusters of activations. It is found to be more robust than voxel-wise comparisons and could potentially lead to more statistical power in fMRI-based group studies.

1 Introduction

Theoretical considerations as well as ample experimental evidence suggest that brain function is, to some extent, supported by the correlated activity of *groups* of neurons. These groups, in turn, tend to be spatially contiguous and consistent with the spatial continuity of anatomical patterns of connectivity and neuronal identity. However, the field of fMRI image analysis is dominated by techniques that use a voxel-wise linear model, the General Linear Model (GLM) approach. As evidence, Grinband *et al.* [1] identified 170 papers published in leading journals just in the first six months of 2007 that used this approach.

An open problem in the fMRI analysis literature is how to appropriately conduct inter-subject studies and summarize their results. The typical approach is to apply the GLM method, which results in candidate regions responsive to the particular experimental stimulus or protocol. This is followed by voxel-wise comparison amongst different subjects. A major drawback of this approach is that voxel-based methods do not give robust inter-subject results. By robustness we mean the sensitivity of the results with respect to small changes in the positions of the voxels or their contents. Several reasons contribute to the lack of robustness across subjects, and even across sessions for the same subjects: misalignment, movement, field distortions and noise, and morphological differences in individual brains. Usually, these problems are addressed by the application of significant spatial smoothing, which effectively leads to a loss of resolution.

A number of recent publications have proposed different algorithms to identify functional clusters. Thirion *et el.* [2, 3] have explored the issue of comparing inter-subject fMRI activation maps. They developed a *functional* parcellation technique based on spectral clustering, to group similar regions across multiple

G.-Z. Yang et al. (Eds.): MICCAI 2009, Part I, LNCS 5761, pp. 1018–1025, 2009.

subjects. Most other methods are heavily based on a priori knowledge of the anatomy and connectivity, in particular sulcal identification [4,5,6]. This peak of interest highlights the need for a consistent measure to compare cluster-based maps.

At the same time, recently developed techniques for functional network analysis have shown promise. They are based on extracting graph structures from spatio-temporal datasets [7], and then computing relevant statistical properties of these graphs. For instance, hub density maps can be computed, which capture the degree of linkage of voxel nodes in the graph. This measure of activity tends to be more highly localized, or compact, than that of the more diffuse activity captured via GLM maps.

The above observations lead naturally to the concept of *resolution*, to describe how well localized the measured brain activity is. We consider a higher resolution technique to be one that potentially generates a larger number of spatially compact candidate regions. Hence there is a need for an improved method to perform group studies and compare maps across subjects. With this aim in mind, we have created the following set of desired capabilities that we seek, and which would be present in an ideal method.

1. Robustness with respect to misregistration. An ideal method is one whose results do not degrade with small spatial shifts in the data, especially when they arise from differences in the brain anatomies of subjects.
2. Robustness with respect to different runs for the same subject. This issue addresses temporal shifts, including temporal partitioning of the data sets, i.e. the analysis of different time segments should yield consistent results if the experimental paradigm is time-invariant.
3. Robustness across multiple scales of resolution. An ideal method would be able to perform comparisons between both high and low resolution data.
4. Generic applicability. An ideal method would be applicable across multiple maps, including maps obtained via GLM or other methods.

The main contribution of this paper is to propose a method called the *weighted cluster coverage* method, which is able to satisfy the four requirements that we seek in an ideal method, as identified above. The cluster coverage method overcomes many of the limitations of applying simple voxel-based image difference or correlation metrics that are typically used. The method is described in detail in Section 2.3.

2 Comparison of Statistical Maps

In this section we present three metrics, namely *voxel correlations*, *weighted set overlap* and *weighted cluster coverage* to compare two fMRI maps. The voxel correlation and set overlap metrics arise naturally under the assumptions made in typical voxel-wise group analysis methods. The weighted cluster coverage metric designed by us, uses spatial contiguity information to overcome the limitations of voxel-wise methods.

2.1 Voxel Correlations

In a typical fMRI-based study, a set of maps is computed for each subject using the GLM analysis. Each map in the set represents "brain activations" in a specific experimental condition, suitably transformed into a normalized statistic such as "Z", "t" or an "F" score. These maps are then aligned to a standardized brain atlas (such as the MNI or Talairch atlases). A voxel-wise comparison of the aligned maps of each subject is carried out using various statistical techniques (e.g., ANOVA) to arrive at the final conclusions of the study [8].

An implicit assumption in such an approach is that value represented by each voxel in the aligned map remains the same (modulo the noise) in a given experimental condition for all subjects in a given group. The voxel-correlation metric is designed to evaluate this assumption.

Given a map μ of N voxels, define $Top(\mu, p)$ as the set of Np voxels with the highest values in the map μ. Here, p denotes a percentile. The *voxel correlation* between two maps μ_1 and μ_2 at a percentile p is defined as the Pearson correlation coefficient between μ_1 and μ_2, restricted to the set $S = Top(\mu_1, p) \cap Top(\mu_2, p)$. Intuitively, voxel correlation represents the similarity between the voxel values of the active voxels of the two maps. Note that if $\mu_1 = \mu_2$, then the voxel correlation is 1; if $\mu_1 = -\mu_2$ then voxel correlation is -1 and if μ_1 and μ_2 are independent then the voxel correlation is zero for all values of p[1].

2.2 Weighted Set Overlap

The machine-learning based fMRI data analysis techniques (such as MVPA [9] or [10]) rely upon extracting a set of meaningful "features" that are used to build models of the data. The model and the corresponding features are used to interpret the results of the experiment and derive scientific conclusions.

An implicit assumption in all such models is that each voxel represents the same physical (or physiological) process in different experimental conditions and subjects. The *weighted set overlap* metric is designed to evaluate this assumption.

Define the weight of a map μ with respect to the set S as the sum of the map values of voxels in the set S. Formally $W(\mu, S) = \sum_{i \in S} \mu_i$. The weighted set overlap of two maps μ_1 and μ_2 at percentile p is given by

$$\frac{W(\mu_1 + \mu_2, S_1 \cap S_2)}{W(\mu_1, S_1) + W(\mu_2, S_2)}$$

where $S_1 = Top(\mu_1, p)$ and $S_2 = Top(\mu_2, p)$. Intuitively, this measure represents the degree of overlap of voxels found active in the two maps. Note that if $\mu_1 = \mu_2$ then the weighted set overlap is 1 for all the values of p. It is 0, if the top p fraction of the voxels of μ_1 and μ_2 do not have any common voxel.

[1] The choice $S = Top(\mu_1, p) \cup Top(\mu_2, p)$ which gives more voxels, was not used since it introduces an artificial negative bias in the correlations.

2.3 The Cluster Coverage Metric

The cluster coverage metric is designed to overcome the limitations of voxel-wise methods by taking spatial contiguity of voxels into account.

Consider two sets of voxels in space, S_1 and S_2. These could arise from two fMRI trials on two subjects, or from two trials on the same subject, for instance. The voxels in the two sets are first clustered into groups of spatially connected 3-d components (two voxels are said to be connected if they share a common face, i.e. we use 6-connectedness in three dimensions). Let these components be denoted by the sets $\{r_{1j}\}$ and $\{r_{2j}\}$. The *weighted cluster coverage* of S_1 by S_2 is defined as the ratio of the weight of clusters of S_1 that intersect with clusters of S_2, to the total weight of clusters of S_1. Formally,

$$C_w(S_1, \mu_1, S_2, \mu_2) = \frac{\sum_{j:r_{1j} \cap S_2 \neq \phi} W(\mu_1, r_{1j})}{\sum_j W(\mu_1, r_{1j})} \tag{1}$$

The weighted cluster coverage $C_w(S_1, \mu_1, S_2, \mu_2)$ is an asymmetric measure, which ranges from 0 to 1. If $C_w(S_1, \mu_1, S_2, \mu_2)$ is zero, this implies that S_1 and S_2 have no voxels in common. As $C_w(S_1, \mu_1, S_2, \mu_2) \rightarrow 1$, this indicates that many connected components of S_1 intersect with connected components in S_2. A cluster coverage of one implies that every spatially contiguous component in S_1 intersects with a component in S_2.

The mean weighted cluster coverage of two maps μ_1 and μ_2 at a percentile p is defined as $(C_w(S_1, \mu_1, S_2, \mu_2) + C_w(S_2, \mu_2, S_1, \mu_1))/2$ where $S_1 = Top(\mu_1, p)$ and $S_2 = Top(\mu_2, p)$. This is a symmetric variant of the cluster coverage metric.

3 Evaluation and Results

3.1 Evaluation Methodology

We used fMRI data from a simple finger tapping experiment. The data consists of fMRI scan of three sessions of six healthy subjects. The sessions, which lasted for 800 seconds, consisted of blocks of a self-paced finger-tapping task. Each session was split into two sub-sessions corresponding to the first and the second half of the session. For each of the sub-sessions, GLM analysis using FSL [11] was carried out to find the areas in the brain active during the finger-tapping task. The resulting maps of Z-statistics were registered to the MNI atlas, and then used for comparisons.

Three types of comparisons were carried out. The *intra-session* comparisons refer to the comparisons between the first half and the second half of the sessions (leading to 6×3 comparisons). The *inter-session* comparisons refer to comparisons between two different sessions of the same subject (leading to $6 \times {}^3C_2 \times 2$ comparisons. Here ${}^nC_r$ refers to the number of ways in which r objects can be chosen from n distinct objects). The *inter-subject* comparisons refer to the comparisons between the same sub-sessions of different subjects (leading to ${}^6C_2 \times 3 \times 2$ comparisons). For each of the comparison, seven different values of

p ranging from 0.2 to 0.002 were considered. All the three comparison metrics (discussed earlier) were computed for all the p-values considered. For each value of p, the mean and standard deviation of the similarity metrics were computed.

3.2 Results

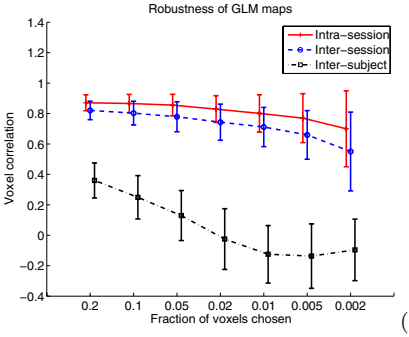
Figure 1(A) shows the voxel correlations of the Z-maps of the finger tapping experiment at different values of p. The mean value of intra-session, inter-session and inter-subject overlap (using the voxel correlation metric) are plotted as a function of p. The error bars represent the standard deviation of the metric.

The mean voxel correlations for inter and intra-session comparison remain in the range 0.87 to 0.70 and 0.85 to 0.55 respectively. The intra-session overlap is slightly higher and within 27 percent of the inter-session overlap for all values of p. In contrast, the mean inter-subject overlap starts at 0.36 and becomes negative for values of p less than or equal to 0.02. This shows that the actual values of the Z-statistic, although fairly consistent within a subject, are highly inconsistent across multiple subjects.

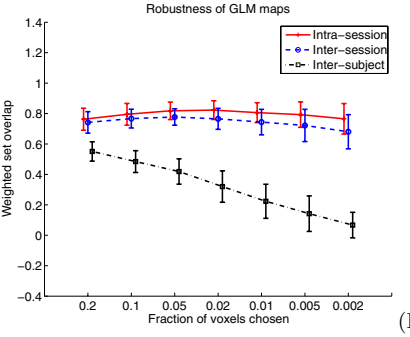
Figure 1(B) shows the weighted set overlap of the Z-maps for different values of p. The results are similar to those of the voxel-correlation metric. For the intra and inter-session comparisons the mean value of weighted set overlap is in the range 0.82 to 0.76 and 0.78 to 0.68 respectively, with inter-session overlap slightly higher than the intra-session overlap (but within 12% of each other). However, the weighted set overlap in the inter-subject comparisons starts at 0.55 for $p = 0.2$ and becomes as small as 0.067 for $p = 0.002$, indicating its inconsistency across the subjects.

The mean weighted coverage measure performs much better as shown in Figure 1(C). For inter- and intra-session comparisons, the mean weighted coverage is almost the same (within 5%) with a low standard deviation. For inter-subject comparison, it is in the range 0.97 to 0.71. This indicates that the weighted cluster coverage measure, which takes into account the spatial contiguity of active voxels, is much more robust for intra- and inter-subject comparisons as compared to the voxel-wise methods discussed above.

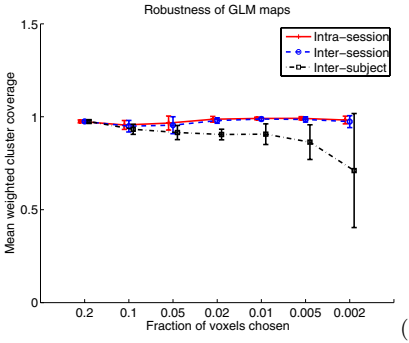
Overlap with misregistration: In order to evaluate the impact of misregistration, the second sub-session of each session was spatially shifted by different amounts. For each shift, all the three overlap measures were computed for $p = 0.05$. Figure 1(D) shows the overlap as a function of the shift amount. The voxel correlations and weighted set overlap are very sensitive to misregistration. A shift of less than 5mm (which is not uncommon across subjects) reduces the voxel correlation from 0.63 to 0.23 and weighted set cover from 0.70 to 0.40. The weighted cluster coverage metric is more tolerant to shifts up to 25mm (primarily because the areas functionally active in the finger tapping task have a similar spatial extent). If the amount of shift is more than 25mm, then all the metrics fall to zero as there is not overlap in the functionally active areas of the first sub-session with the misregistered second sub-session.



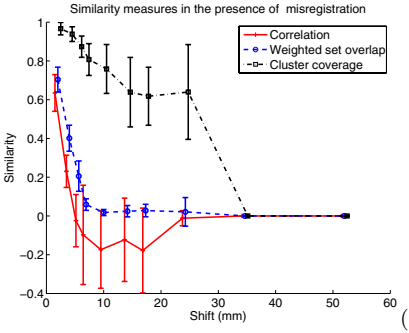
(A)



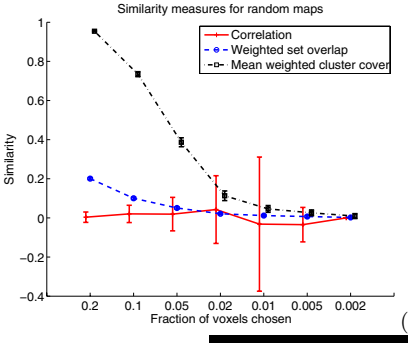
(B)



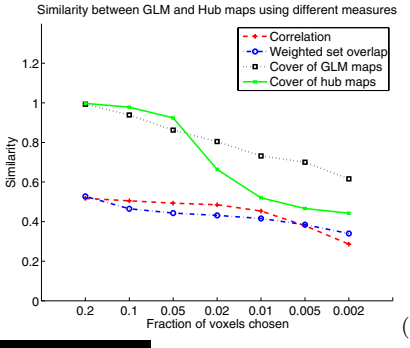
(C)



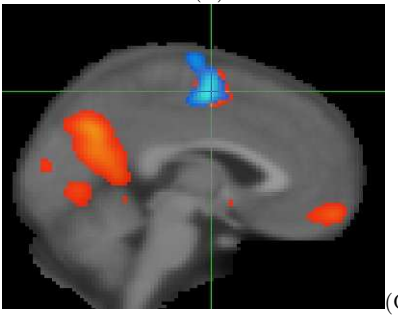
(D)



(E)



(F)



(G)

Fig. 1.

Random Maps: The weighted cluster coverage metric has a positive bias since it computes the overlap of spatially contiguous clusters. If a large fraction of voxels are chosen, many clusters of the one map may intersect with the second map by chance, leading to an unreasonably high weighted cluster coverage value. To evaluate the extent of this bias, we compared randomly generated maps using the three metrics.

To make a random map, first each voxel was assigned a normally distributed random value. To make the map similar to fMRI activation maps, a spatially smoothing 2.5mm Gaussian filter was applied to the randomly generated voxel values. The resulting map was masked using the brain mask of one of the subject in the study. Now pairs of random maps were compared with each other.

Figure 1(E) shows the overlap of random maps. The voxel correlations remain close to zero for all the values of p, as expected. The bias in weighted set overlap follows the p-value as expected. The weighted cluster coverage shows biases of $0.95, 0.73, 0.39$ for values of p in $0.2, 0.1, 0.05$ respectively. However, for $p = 0.02, 0.01, 0.005, 0.002$, this bias rapidly diminishes to $0.1, 0.05, 0.02, 0.01$ respectively. These results are not unexpected since the spatially contiguous clusters become large when large fraction of voxels is chosen. This increases the chance of intersection of clusters, making the weighted cluster coverage metric more liberal at large values of p. However, when the fraction of voxels chosen is small (say ≤ 0.02), this bias quickly reduces to acceptably small values.

In general, random maps induce a probability distribution on the value of the weighted cover metric. This distribution may be used to convert the cover metric to a p-value representing the chance likelihood of the event.

Maps generated using different techniques: Brain maps can be generated using several other techniques in addition to GLM. For example the network analysis technique [7] generates *hub maps* representing the density of functional network connections. The functionally active areas are also found to have high connectivity and therefore high link-density in the hub maps. In addition, several other areas of high link-density which are not found active in the GLM analysis are also present in the hub maps. Figure 1(G) shows one such slice of a subject in the finger tapping experiment, where the threshold was held at the same percentile for both maps. The Supplementary Motor Area, which is found active by GLM analysis (colored blue), also has high link-density (red). However, one can see other areas that have high link-density (red) but not found active by the GLM analysis; in particular, there is a prominent cluster centered in the Posterior Cingulate that is not locked to the experimental protocol.

Figure 1(F) shows the overlap of GLM maps and the hub maps using different metrics. The overlap using voxel correlations and the weight set cover is symmetric and in the range 0.5 to 0.3. However, the cluster coverage metric shows much larger overlap starting from close to 1 for high values of p to close to 0.6 for low values of p. The asymmetric nature of the cluster coverage measure also uncovers the property of these maps discussed above, also visible in Figure 1(G). This confirms that most of the clusters found active by GLM analysis also contain voxels with high connectivity, and that the converse is not true.

4 Conclusions and Future Work

We have shown that the cluster coverage metric gives better consistency, lower standard deviations and robustness with respect to activation thresholds than metrics that do not exploit the a priori information about the spatial contiguity of brain function. The metric clearly reduces intra-subject variability, but its effect is more dramatic for inter-subject comparisons, making it particularly promising for large group studies. The effectiveness of the metric is also reflected in the slow degradation upon misregistration, when compared to non-cluster metrics, a feature of particular importance for both inter- and intra-subjects studies. Our metric meets the four requirements for an ideal comparison method, as described in the introduction.

This metric could be extended to the surface-based methods [5] by computing spatially contiguous clusters (and their intersections) on the manifold defined by the cortical surface.

References

1. Grinband, J., Wager, T.D., Lindquist, M., Ferrera, V.P., Hirsch, J.: Detection of time-varying signals in event-related fMRI designs. Neuroimage 43(3), 509–520 (2008)
2. Thirion, B., Flandin, G., Pinel, P., Roche, A., Ciuciu, P., Poline, J.B.: Dealing with the shortcomings of spatial normalization: Multi-subject parcellation of fMRI datasets. Hum. Brain Mapping 27(8), 678–693 (2006)
3. Vincenf, T., Ciuciu, P., Thirion, B.: Sensitivity analysis of parcellation in the joint detection-estimation of brain activity in fMRI. In: 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 568–571 (2008)
4. Meyer, J.W., Makris, N., Bates, J.F., Caviness, V.S., Kennedy, D.N.: MRI-based topographic parcellation of human cerebral white matter: 1. Technical foundations. Neuroimage 9(1), 1–17 (1999)
5. Lohmann, G., von Cramon, D.Y.: Automatic labelling of the human cortical surface using sulcal basins. Medical Image Analysis 4(3), 179–188 (2000)
6. Cachia, A., Mangin, J.F., Rivière, D., Papadopoulos-Orfanos, D., Kherif, F., Bloch, I., Régis, J.: A generic framework for parcellation of the cortical surface into gyri using geodesic voronoi diagrams. Medical Image Analysis 7(4), 403–416 (2003)
7. Eguíluz, V.M., Chialvo, D.R., Cecchi, G.A., Baliki, M., Apkarian, V.A.: Scale-free brain functional networks. Physical Review Letters 94, 018102 (2005)
8. Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D.: Statistical Parametric Mapping, The Analysis of Functional Brain Images. Academic Press, London (2007)
9. Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V.: Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. Trends in Cognitive Sciences 10(9), 424–430 (2006)
10. Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K., Malave, V.L., Mason, R.A., Just, M.A.: Predicting human brain activity associated with the meanings of nouns. Science 320(5880), 1191–1195 (2008)
11. FSL Release 3.3, http://www.fmrib.ox.ac.uk/fsl

Author Index

Abbott, Jake J. I-540
Abd-Elmoniem, Khaled Z. II-331
Abolmaesumi, Purang I-75, I-803, II-795
Abou El-Ghar, Mohamed II-682, II-943
Abugharbieh, Rafeef II-474, II-490
Adams, Judith E. II-1017
Aganj, Iman II-423
Aguet, François II-625
Ahmed, Hashim Uddin I-787
Aickelin, Uwe I-984
Aja-Fernández, Santiago I-919, I-951, II-415
Akselrod-Ballin, Ayelet I-632
Al-Ahmad, Amin I-9
Alberola-López, Carlos I-156
Albrecht, Thomas II-1
Aldridge, Ben II-1100
Alexander, Daniel C. I-329
Allain, Baptiste I-491
Allard, Jérémie I-198
Allen, Clare I-787
Alvino, Christopher II-34
Amunts, Katrin II-795
An, Hongyu II-232
Ananthanarayanan, Rajagopal I-861
Andersson, Mats I-1000
Angelini, Elsa D. I-222, II-1025
Anquez, Jérémie II-1025
Anwander, Alfred I-886
Arai, Andrew II-741
Arbel, Tal II-109
Argibay-Quiñones, Daniel I-903
Armspach, Jean-Paul I-959, II-566
Arnold, Douglas L. II-584, II-601
Arzhaeva, Yulia II-724
Ashraf, Haseem II-699
Assemlal, Haz-Edine II-406
Atasoy, Selen I-499
Athanasiou, Thanos I-410
Atkins, M. Stella II-1108
Auzias, Guillaume I-730
Avedissian, Christina II-506
Avramescu, Taina II-316
Awate, Suyash P. II-141, II-200, II-852
Ayache, Nicholas I-214, I-297

Baillet, Sylvain I-730
Bainbridge, Daniel I-516
Balicki, Marcin I-108
Ballerini, Lucia II-1100
Barenstz, Jelle O. II-836
Barillot, Christian II-584
Barkovich, A. James I-289
Barmpoutis, Angelos I-640
Barratt, Dean I-787
Barrot Cortés, Emilia II-275
Barysheva, Marina I-845, I-967, II-498, II-506
Basser, Peter J. I-181
Baumann, Michael I-67
Bazin, Pierre-Louis I-943
Becher, Harald II-893
Beg, Mirza Faisal I-713, II-549
Bello, Fernando I-34
Ben Ayed, Ismail II-373, II-901
Bergeles, Christos I-540
Bernhardt, Boris I-214
Beširević, Armin II-356
Beymer, David I-648
Bhotika, Rahul I-239
Bi, Jinbo II-1009
Bibin, Lazar II-1025
Biesdorf, Andreas I-607
Bijnens, Bart H. I-140, II-759
Biros, George II-257, II-885
Bischof, Horst II-860
Blake, Andrew II-558
Blanc, Julie I-214
Blanc, Rémi II-84
Bloch, Isabelle I-222, II-1025
Blumberg, Hilary P. II-18
Bock, Davi I-632
Boctor, Emad I-507
Boesecke, Robert I-402
Boettger, Thomas I-828
Bogovic, John I-943
Boisgontier, Hervé I-959

Boissonnat, Jean-Daniel II-283
Boisvert, Jonathan I-803
Boltcheva, Dobrina II-283
Boselie, Toon I-524
Bossa, Matías Nicolás II-240
Boubekeur, Tamy II-1025
Boudjemline, Younes I-214
Bouix, Sylvain I-345
Brady, Michael II-398
Breteler, Monique M.B. I-853
Brodlie, Ken W. II-43
Brooks, Rupert I-51
Brost, Alexander I-394
Brückner, Marcel I-386
Brun, Caroline C. II-498, I-967
Brun, Luc II-406
Buatti, John II-827
Buechert, Martin II-665
Buhmann, Joachim M. II-633
Bullitt, Elizabeth I-321
Bulpitt, Andy J. II-43
Bunyak, Filiz II-617
Burdette, Everette C. I-59

Caan, Matthan I-869
Cahill, Nathan D. I-574
Cairncross, J. Gregory II-522
Camus, Estelle I-9
Cao, Kunlin I-739
Capuano, Ermanno I-369
Cárdenes, Rubén I-903
Cardoso, M. Christina II-617
Cardoso, Manuel Jorge II-441
Carlier, Stéphane II-776
Carneiro, Gustavo II-575
Carney, Paul R. I-190
Carr, Peter II-976
Carswell, Melody C. II-1067
Casula, Roberto I-34
Cates, Joshua II-167
Cavallaro, Alexander I-255
Cecchi, Guillermo A. I-1018
Chagin, Vadim II-617
Chan, Kap Luk II-803
Chan, Tony F. I-337, II-133
Chandran, Sharat I-230
Chang, Yu-Bing II-968
Chappell, Michael A. II-514
Chefdhotel, Christophe II-741
Chen, Albert Y. II-715
Chen, Elvis I-803
Chen, Mei I-100
Chen, Terrence II-877
Chen, Ting I-43
Chen, Yasheng II-232
Cheng, Guang I-190
Cheng, Jian I-911
Chiang, Ming-Chang I-967, II-506
Chiba, Toshio I-83
Chiu, Ing-Sh II-266
Chng, Nick II-960
Chong, Jaron II-373, II-901
Chou, Yi-Yu I-967, II-498
Christensen, Gary E. I-739, II-795
Chuang, Ming II-100
Chung, Albert C.S. I-696, II-59
Chung, François II-1051
Chung, Moo K. II-158, II-549
Ciompi, Francesco II-869
Ciuciu, Philippe I-975
Clark, Chris A. II-150
Clark, James I-410
Clarke, Colin I-516
Clarke, Duncan II-1067
Clarkson, Matthew J. II-441
Clarysse, Patrick II-365
Clayden, Jonathan D. II-150
Cobzas, Dana II-531
Cohen, Laurent II-1042
Collins, D. Louis II-109, II-584, II-592, II-601
Colliot, Olivier I-730
Comaniciu, Dorin I-17, I-116, I-255, II-575, II-767, II-877
Combès, Benoît II-175
Commowick, Olivier II-993
Connelly, Kim II-750
Cook, Richard I-491
Cootes, Tim F. II-1017
Corso, Jason II-715
Cotin, Stéphane I-198, II-291, I-377
Coukos, George I-230
Coupé, Pierrick II-601
Couton, Mathieu II-300
Criminisi, Antonio II-558

D'Haese, Pierre-François I-557
Daanen, Vincent I-67
Dahnke, Robert II-125

Dai, Wei II-133
Dalton, Kim M. II-158
Darzi, Ara I-410
Das, Sandhitsu R. II-141
Dassopoulos, Themos I-582
Datar, Manasi II-167
Davatzikos, Christos I-680, II-257
Davidson, Richard J. II-158
Dawant, Benoit M. I-557
De, Suvranu II-348
de Boer, Renske I-853
de Bruijne, Marleen II-51, II-699
De Craene, Mathieu I-140, II-759
de Groot, Marius I-853
Deguchi, Daisuke II-707
de Hoop, Bartjan I-263
Deinzer, Frank I-386, I-590
de Jong, Pim A. II-724
de Korte, Chris L. II-927
Delingette, Hervé I-214, II-1051
del Nido, Pedro J. I-26
Delorme, Sébastien I-51
Deng, Zhigang II-968
Denzler, Joachim I-386
Dequidt, Jeremie I-377
Deriche, Rachid I-911
De Robles, Paula II-522
Descoteaux, Maxime I-886, II-482
Deux, Jean-François I-672
Devlin, Steven I-75
Dewan, Maneesh II-715, II-1033
de Zubicaray, Greig I. I-845, I-967, II-498, II-506
Diallo, Mamadou I-9, I-828
Díaz, Alejandro II-690
Dick, Alexander II-750
Ding, Kai I-739
Dinov, Ivo II-208
Dione, Donald P. I-688
Doblaré Castellano, Manuel II-275
Dobrucki, Lawrence W. I-688
Dohi, Takeyoshi I-83, I-418, I-451
Dojat, Michel II-540
Donner, René II-860
Dougherty, Robert F. I-861
Drabycz, Sylvia II-522
Dubuisson, Florian I-475
Duchateau, Nicolas II-759
Duncan, James S. I-206
Dundar, Murat II-1009
Duong, Christophe I-9
Duriez, Christian I-377, II-291
Durrleman, Stanley I-214, I-297

Ebrahimi, Mehran I-811
Edenbrandt, Lars I-664
Edwards, Philip J. I-34
Eggers, Georg I-402
Ehrhardt, Jan I-755
Eklund, Anders I-1000
El-Baz, Ayman I-281, II-682, II-943
Elhawary, Haytham I-837
Elson, Daniel I-483
Emberton, Mark I-787
Engelbrecht, Rainer I-467
Eom, Jaesung II-348
Ernst, Floris II-356
Ersoy, Ilker II-617
Essafi, Salma II-919
Estépar, Raúl San José II-690
Euler, Ekkehard I-173

Fahmy, Ahmed S. II-985
Fahrig, Rebecca I-9
Falk, Robert II-682, II-943
Fang, Le I-198
Faraco, Carlos I-313
Fei, Jin II-1084
Feldman, Michael D. I-230
Fernández-Nofrerías, Eduard II-869
Feulner, Johannes I-255
Feußner, Hubertus I-459, I-467
Fichtinger, Gabor I-59, I-803
Figl, Michael I-34
Fillard, Pierre I-886, I-927
Fischl, Bruce I-598, II-1075
Fisher, Robert II-1100
Fleming, Ioana I-507
Fletcher, P. Thomas II-167
Fleuret, François II-625
Floros, Xenofon II-633
Fonteijn, Hubert I-329
Forbes, Florence II-540
Foroughi, Pezhman I-507
Fox, Nick C. II-441
Frangi, Alejandro F. I-140, II-759
Fua, Pascal II-625
Fuchs, Thomas J. II-633
Funka-Lea, Gareth I-222, II-34

Gangl, Alfred I-247
García-Lorenzo, Daniel II-584
Garg, Rahul I-1018
Gaser, Christian II-125
Gedamu, Elias II-601
Gee, James C. II-141, II-200, II-466, II-852
Gehlbach, Peter I-108
Gelas, Arnaud II-641
Geng, Xiujuan II-795
Georgescu, Bogdan I-17, II-767
Gerber, Samuel I-305
Gerig, Guido I-297, I-321, II-167
Ghosh, Aurobrata I-911
Giannarou, Stamatia I-483, I-499
Gibson, Eli I-713
Gill, Sean I-803
Gilmore, John II-232
Gimel'farb, Georgy I-281, II-682, II-943
Ginel Cañamaque, Ángel II-275
Girgis, Hani I-582
Glaunès, Joan I-730
Glenn, Orit A. I-289
Glocker, Ben I-499, I-672
Goh, Alvina I-877
Goksel, Orcun II-248, II-960
Golland, Polina I-272, I-565, I-598, I-1009, II-1075
Gong, Ren Hui I-75
González Ballester, Miguel A. II-275
González, Germán II-625
Goodlett, Casey I-321
Gouaillard, Alexandre II-641
Gouttard, Sylvain I-321, II-167
Grady, Leo II-910
Grau, Vicente II-893
Grimbergen, Cornelis I-869
Grisoni, Laurent II-291
Grossman, Murray II-852
Gruionu, Gabriel II-316
Gruionu, Lucian Gheorghe II-316
Gu, Xianfeng II-133
Guébert, Christophe II-291
Guehring, Jens II-741, II-910
Guevara, Pamela I-935
Guillon, Jean-Pierre II-976
Guo, Lei I-313, II-184, II-458
Gurmeric, Serhan II-776
Gutiérrez Boronat, Javier I-459
Habas, Piotr A. I-289
Hadjidemetriou, Stathis II-665
Häfner, Michael I-247
Hager, Gregory D. I-91, I-426, I-435 I-507, I-582
Hall, Matt G. I-329
Hämäläinen, Matti I-1009
Hamarneh, Ghassan II-649
Hambrock, Thomas II-836
Hamm, Jihun I-680
Han, Jae-Ho I-108
Han, Qiong II-1067
Handa, James I-108
Handels, Heinz I-755
Harel, Noam II-423
Hartley, Richard II-976
Hartov, Alexander I-795, II-308
Hasegawa, Yosihnori II-707
Hata, Nobuhiko I-1, I-837
Hawkes, David J. I-34, I-491, I-574, I-787
Heining, Sandro Michael I-173
Heitz, Fabrice I-959, II-566
Hennig, Juergen II-665
Heredia, David II-682, II-943
Herrero Jover, Javier II-275
Hervé, Pierre-Yves I-984
Hietala, Jarmo II-216
Hinrichs, Chris II-786
Ho, Harvey II-323
Ho, Hon Pong II-18
Hoffman, Eric A. I-739
Hogeweg, Laurens II-724
Höller, Kurt I-459, I-467
Horiuchi, Tetsuya I-451
Hornegger, Joachim I-17, I-132, I-255, I-394, I-459, I-467, I-549, II-26, II-575, II-767, II-819
Houle, Helene II-767
Howe, Robert D. I-26
Hu, Mingxing I-34, I-491
Hu, Yipeng I-787
Huang, Albert II-474
Huang, Wei II-803
Huang, Xiaolei II-673, II-1059
Huber, Martin I-17, II-575
Huisman, Henkjan J. II-836, II-927
Hunter, Peter II-323
Hurtig, Mark I-75

Ibrahim, Joseph G. II-192
Idier, Jérôme I-975
Iliescu, Nicolae II-316
Inoue, Jiro I-75
Ionasec, Razvan Ioan I-17, II-767
Iordachita, Iulian I-108
Irfanoglu, Mustafa Okan I-181
Iseki, Hiroshi I-443
Isguder, Gozde Gul II-776
Ishii, Masaru I-91
Ishikawa, Hiroshi I-100
Islam, Ali II-373, II-901
Itoh, Kazuko I-443
Iwano, Shingo II-707

Jacques, Robert II-100
Jacques, Steven L. II-657
Jäger, Florian II-819
Jagersand, Martin II-531
Jahanshad, Neda II-498
Jahn, Jasper I-459
Jain, Ameet K. I-59
Janka, Rolf II-819
Janowczyk, Andrew I-230
Jayender, Jagadeesan I-1
Jenkinson, Mark I-705, II-951
Jezzard, Peter II-514
Ji, Songbai I-795, II-308
Ji, Yongnan I-984
Jiang, Di I-51
Jiang, Tianzi I-911
John, Matthias I-9
Johnson, Sterling II-786
Johnsrude, Ingrid S. II-795
Joldes, Grand Roman II-300
Jolly, Marie-Pierre II-910
Joshi, Sarang I-305

Kabus, Sven I-747
Kadir, Timor I-771, II-34
Kadoury, Samuel II-92
Kagiyama, Yoshiyuki I-532
Kainmueller, Dagmar II-76
Kaiser, Hans-Jürgen I-607
Kakadiaris, Ioannis A. II-885
Kane, Gavin I-402
Kang, Jin I-108
Kao, Chris I-557
Kapusta, Livia II-927
Kauffmann, Claude I-475
Kazhdan, Michael II-100
Keil, Andreas II-389
Keller, Merlin II-450
Kellman, Peter II-741
Keriven, Renaud II-482
Kerrien, Erwan I-377
Khamene, Ali I-9, I-828, II-381
Khan, Ali R. I-713, II-549
Khudanpur, Sanjeev I-426
Kier, Christian II-935
Kikinis, Ron II-690
Kim, Kio I-289
Kim, Peter T. II-158
Kim, Sung I-43
Kindlmann, Gordon I-345
King, Martin D. II-150
Kirchberg, Klaus J. II-68
Kitasaka, Takayuki II-707
Kleemann, Markus II-356
Klein, Stefan I-369, I-853
Klinder, Tobias I-747
Knösche, T.R. I-886
Knutsson, Hans I-1000
Koay, Cheng Guan I-181
Kobayashi, Kazuto I-532
Koch, Christoph II-356
Kohlberger, Timo II-34
Komodakis, Nikos I-672
Konrad, Peter E. I-557
Krishnan, Arun II-715
Kumar, Rajesh I-582
Kumar, Vedant II-943
Kunz, Manuela I-75
Kurhanewicz, John II-844
Kutter, Oliver I-763
Kwitt, Roland I-247
Kwok, Ka-Wai I-410
Kybic, Jan II-365

Labat, Christian I-59
Lahrech, Abdelkabir I-148
Lamecker, Hans II-76
Langs, Georg II-860, II-919
Lara Rodríguez, Laura II-275
Lauritsch, Günter II-68, I-132, II-389
Lavielle, Marc II-450
Law, Max W.K. II-59
Le, Binh Huy II-968
Lecoeur, Jeremy II-584
Lee, Agatha D. I-967, II-498

Lee, Junghoon I-59
Lee, Tim K. II-1108
Leemans, Alexander I-853
Legg, Philip A. I-616
Lenglet, Christophe I-877, II-423
Leow, Alex D. I-845
Leow, Wee Kheng II-266
Leporé, Natasha I-967, II-498
Lerotic, Mirna I-410
Lesage, David I-222
Levner, Ilya II-522
Li, Chao I-165
Li, Fuhai II-609
Li, Gang I-313, II-184, II-458
Li, Hao II-266
Li, Hongsheng II-673
Li, Hua II-1042
Li, Huiqi II-803
Li, Kaiming I-313
Li, Rui I-557
Li, Shuo II-373, II-901
Li, Xiang II-1100
Li, Yang I-624
Li, Zhixi II-1001
Liang, Jianming II-1009
Liao, Hongen I-83, I-418, I-451
Liao, Rui I-394
Liao, Shu I-696
Licht, Daniel II-200
Lieberman, Jeffrey II-192
Lim, Joo Hwee II-803
Lin, Ching-Long I-739
Lin, Fa-Hsuan I-1009
Lin, Henry I-426
Lin, Weili I-321, I-721, II-192, II-232
Lin, Zhuohua I-443
Ling, Haibin I-116
Linguraru, Marius George II-1001
Linte, Cristian A. I-361
Liu, Fenghong II-308
Liu, Huafeng II-732
Liu, Jiang II-803
Liu, Tianming I-313, II-184, II-458
Liu, Xiaofeng II-331
Liu, Yunlong II-827
Lo, Pechin II-51, II-699
Lollis, S. Scott II-308
Lopata, Richard G.P. II-927
López Villalobos, José Luís II-275
Lorenz, Christine H. II-741
Lorenz, Cristian I-747
Lovat, Laurence B. I-491
Lu, Le II-715, II-1009
Lu, Yingli II-750
Lurz, Philipp I-214
Lüthi, Marcel II-1
Lythgoe, Mark F. I-329

Ma, Songde I-198
Machiraju, Raghu I-181
Madabhushi, Anant I-230, II-844
Madsen, Sarah K. I-967
Mahdavi, Seyedeh Sara II-339, II-960
Mahfouz, Ahmed Essam II-985
Makrogiannis, Sokratis I-239
Malandain, Grégoire II-993
Malcolm, James G. I-894
Mallinson, Gordon II-323
Mangin, Jean-François I-730, I-927, I-935, II-117
Manjón, José V. II-601
Mansi, Tommaso I-214
Marchal, Maud II-291
Mareci, Thomas H. I-190
Mariano-Goulart, Denis I-148
Marmulla, Rüdiger I-402
Marshall, David I-616
Martín-Fernández, Marcos I-156, I-903
Martín-Fernández, Miguel Ángel I-156
Martel, Anne L. I-811
Martens, Volker II-356
Masamune, Ken I-83, I-418, I-451
Mateus, Diana I-499
Mauri, Josepa II-869
McGregor, Robert H.P. I-124
McKeown, Martin J. II-490
McLaughlin, Robert A. II-657
McLennan, Andrew II-398
McMahon, Katie L. I-845, I-967, II-498, II-506
McNutt, Todd II-100
Megason, Sean II-641
Meining, Alexander I-499
Melhem, Elias R. II-141
Melvær, Eivind Lyche I-771
Mendizabal-Ruiz, E. Gerardo II-885
Menke, Ricarda A. I-705
Mertins, Alfred II-935
Metaxas, Dimitris I-43
Metz, Coert T. I-369

Meyer-Wiethe, Karsten II-935
Michaud, Gregory F. I-1
Milko, Sergiy I-771
Miller, James V. I-239
Miller, Karol II-300
Miller, Stephen I-313
Mirota, Daniel I-91
Mitchell, Ross II-522
Mithraratne, Kumar II-323
Mizutani, Masahiro I-451
Moch, Holger II-633
Modat, Marc II-441
Modha, Dharmendra S. I-861
Moore, John I-361, I-516
Mora, Vincent I-51
Moradi, Mehdi II-339
Morgan, James E. I-616
Mori, Kensaku II-707
Mori, Masaki II-707
Morosan, Patricia II-795
Morra, Jonathan H. II-432
Morris, William J. II-339, II-960
Morrison, Paul R. I-837
Mosaliganti, Kishore II-641
Mosayebi, Parisa II-531
Motomura, Noboru I-451
Motreff, Pascal I-475
Mountney, Peter I-483
Mousavi, Parvin I-803
Mozer, Pierre I-67
Mühling, Joachim I-402
Mullin, Gerard I-582
Muñoz-Moreno, Emma I-903
Muralidhar, Krishnamurthy I-124
Murphy, Keelin I-747
Murtha, Albert II-531
Murthy, Jayasimha II-1084
Mylonas, George P. I-353, I-410

Nakamoto, Masahiko I-532
Natori, Hiroshi II-707
Navab, Nassir I-173, I-499, I-763, I-779, II-389, II-767
Neefjes, Lisan A. I-369
Neimat, Joseph S. I-557
Nelson, Bradley J. I-540
Ng, Bernard II-490
Nichols, Thomas E. I-992
Nie, Jingxin I-313, II-184, II-458
Nielles-Vallespin, Sonia II-741
Nielsen, Mads II-699
Niessen, Wiro J. I-369, I-853
Nillesen, Maartje M. II-927
Noble, J. Alison I-574, II-893
Noblet, Vincent I-959, II-566
Noche, Ramil II-641

Obholzer, Nikolaus II-641
Oguro, Sota I-837
Oh, Teik II-1017
Ohlsson, Henrik I-1000
Ohlsson, Mattias I-664
Okada, Toshiyuki I-532, II-811
Okell, Thomas W. II-514
Olmos, Salvador II-240
Ostermeier, Martin I-116, II-877
Ota, Shunsuke II-707
Otomaru, Itaru I-532
Ou, Wanmei I-1009
Ourselin, Sebastien I-491, II-441

Pace, Danielle I-516
Pacheco, Elisa II-1017
Pajevic, Sinisa I-181
Palaniappan, Kannappan II-617
Pallavaram, Srivatsan I-557
Pan, Chunhong I-198
Panagiotaki, Eleftheria I-329
Papademetris, Xenophon I-206, I-688, II-18
Paragios, Nikos I-672, II-92, II-919
Pastrama, Stefan II-316
Patel, Rajni V. I-1
Paulsen, Keith D. I-795, II-308
Pavlidis, Ioannis T. II-1084, II-1092
Péchaud, Mickaël II-482
Pedersen, Jesper Johannes Holst II-51
Pendsé, Doug I-787
Penne, Jochen I-459, I-467, I-549
Pennec, Xavier I-214, I-297
Penney, Graeme P. I-34
Pérez, Frederic II-275
Pérez del Palomar, Amaya II-275
Perrin, Douglas P. I-26
Perrot, Matthieu I-730, II-117, II-450
Peters, Terry M. I-361, I-516
Petersen, Dirk II-356
Petrović, Aleksandar I-705
Pfefferbaum, Adolf II-224
Pham, Dzung L. I-943

Pichora, David I-803
Pitiot, Alain I-984
Platel, Bram I-524
Pluim, Josien P.W. I-263, I-747
Poupon, Cyril I-886, I-927, I-935
Poynton, Clare II-951
Pozo, Jose M. I-140
Prastawa, Marcel I-321
Price, Anthony I-329
Prima, Sylvain II-175
Prince, Jerry L. I-59, I-943, II-331
Prokop, Mathias I-263
Pruessner, Jens C. II-592
Prümmer, Marcus I-132
Prummer, Simone I-116, II-877
Pujol, Oriol II-869
Punithakumar, Kumaradevan II-373, II-901
Pura, John A. II-1001

Raczkowsky, Jörg I-402
Radau, Perry II-750
Radeva, Petia II-869
Raij, Tommi I-1009
Rainey, Sabrina II-682
Rajagopal, Gunaretnam I-43
Rajan, Purnima I-582
Rajpoot, Kashif II-893
Rangarajan, Anand I-648
Rao, A. Ravishankar I-1018
Rathi, Yogesh I-894
Rechsteiner, Markus P. II-633
Reed, Galen II-844
Rees, Jonathan II-1100
Reich, Daniel I-943
Reichardt, Wilfried II-665
Reid, R. Clay I-632
Reiley, Carol E. I-426, I-435
Reinartz, Rianne I-524
Reinhardt, Joseph M. I-739
Remple, Michael S. I-557
Renard, Félix I-959
Reyes, Mauricio II-84
Ricco, Susanna I-100
Rico, Agnès I-148
Ridgway, Gerard R. II-441
Rietzel, Eike I-828
Riklin Raviv, Tammy I-272
Risser, Laurent I-975
Rivaz, Hassan I-507
Rivière, Denis I-935, II-117
Robbins, Peter II-657
Roberts, David W. I-795, II-308
Roberts, Martin G. II-1017
Robles, Montserrat II-601
Roca, Pauline I-935
Roche, Alexis II-450
Rodríguez Panadero, Francisco II-275
Rohkohl, Christopher I-132, I-549
Rohlfing, Torsten II-224
Rohr, Karl I-607
Roldan, Gloria II-522
Rosen, Mark II-844
Rosin, Paul L. I-616
Ross, Ian II-373
Ross, James C. II-690
Rousseau, François I-289, II-566
Rueckert, Daniel I-34, II-466
Rumbach, Lucien I-959, II-566
Rydell, Joakim I-1000

Sabuncu, Mert R. I-565, I-598, II-1075
Sadeghi, Maryam II-1108
Sage, Caroline I-869
Sahu, Mahua I-787
Salcudean, Septimiu E. II-248, II-339, II-960
Salganicoff, Marcos II-715, II-1009
Salimi-Khorshidi, Gholamreza I-992
Sampson, David D. II-657
Samset, Eigil I-771
Sandberg, Jesse K. II-1001
Sands, Gregory II-323
Sapiro, Guillermo II-423
Sarrut, David II-365
Sarry, Laurent I-475
Sarunic, Marinko II-649
Sasaki, Tomoya I-443
Sasaroli, Dimitra I-230
Sato, Yoshinobu I-532, II-811
Saunders, Christobel II-657
Savadjiev, Peter I-345
Schaap, Michiel I-369
Schaller, Christian I-549
Scheinost, Dustin I-688
Scherrer, Benoit II-540
Schlichting, Stefan II-356
Schmauss, Bernhard I-467
Schmid, Holger II-323
Schmidt-Richberg, Alexander I-755

Schneider, Armin I-459, I-467
Schrauder, Thomas I-467
Schultz, Carl I-369
Schuman, Joel I-100
Schwab, Siegfried II-819
Schweikard, Achim II-356
Scolaro, Loretta II-657
Seales, Williams B. II-1067
Seidel, Günter II-935
Seifert, Sascha I-255
Seiler, Christof II-84
Seim, Heiko II-76
Sermesant, Maxime I-214
Serruys, Patrick W. I-369
Seshamani, Sharmishtaa I-582
Sessa, Salvatore I-443
Shaffer, Teresa II-682
Shamaei, Kamran I-540
Sharma, Swati II-566
Shen, Dinggang I-656, I-721, II-232
Shen, Tian II-673, II-1059
Shenton, Martha E. I-345, I-894
Sherbondy, Anthony J. I-861
Shi, Chengyu II-348
Shi, Pengcheng II-732
Shi, Xiaoyan II-192
Shi, Yonggang II-208
Shotton, Jamie II-558
Shyn, Paul B. I-837
Silva, Etel II-759
Silverman, Edwin K. II-690
Silverman, Stuart G. I-837
Simari, Patricio II-100
Singh, Rajendra I-230
Singh, Vikas II-158, II-786
Sinusas, Albert J. I-206, I-688
Siow, Bernard I-329
Sitges, Marta II-759
Sjöstrand, Karl I-664
Skinner Jr., John I-239
Slosman, Daniel O. II-34
Smith, Benjamin II-649
Smith, Mark II-827
Smith, Stephen M. I-705, I-992
Song, Danny Y. I-59
Song, Qi II-827
Song, Yi II-43
Sonka, Milan II-827
Sørensen, Lauge II-699
Sotiras, Aristeidis I-672
Soza, Grzegorz II-26
Spadinger, Ingrid II-960
Spinas, Giatgen II-633
Sporring, Jon II-51, II-699
Staib, Lawrence H. II-18
Stewart, James I-75
Stippich, Christoph I-607
Stoyanov, Danail I-353
Strauss, Olivier I-148
Strobel, Norbert I-394
Strup, Stephen E. II-1067
Studholme, Colin I-289
Stürmer, Michael I-467, I-549
Styner, Martin II-192
Suenaga, Yasuhito II-707
Sugano, Nobuhiko I-532, II-811
Suh, Jung W. I-688
Sullivan, Edith V. II-224
Summers, Ronald M. II-9, II-1001
Sun, Loi Wah I-410
Sun, Ying I-165
Sun, Zhong Yi II-117
Sunaert, Stefan I-869
Sundar, Hari II-257, II-381
Suzuki, Takashi I-443
Syeda-Mahmood, Tanveer I-648
Szczerba, Dominik I-124
Székely, Gábor I-124, II-84
Szmigielski, Cezary II-893

Tada, Yukio I-532, II-811
Tahmasebi, Amir M. II-795
Takabatake, Hirotsugu II-707
Takamoto, Shinichi I-451
Takanishi, Atsuo I-443
Takao, Masaki I-532, II-811
Tam, Roger II-474
Tao, Yimo II-715
Tasdizen, Tolga I-305
Tatli, Servet I-837
Taylor, Andrew M. I-17, I-214
Taylor, Russell H. I-91, II-100, I-108
Telle, Benoît I-148
ter Haar Romeny, Bart I-524
Thijssen, Johan M. II-927
Thompson, Paul M. I-337, I-845, I-877, I-967, II-133, II-432, II-498, II-506
Tiwari, Pallavi II-844
Toews, Matthew II-109

Toga, Arthur W. I-337, I-845, I-967, II-133, II-208, II-432, II-498, II-506
Tohka, Jussi II-216
Trabelsi, Olfa II-275
Traub, Joerg I-173
Tristán-Vega, Antonio I-919, I-951, II-415
Troccaz, Jocelyne I-67
Trouvé, Alain I-297, I-730
Tschumperlé, David II-406
Tsiamyrtzis, Panagiotis II-1092
Tsin, Yanghai II-68
Tsukihara, Hiroyuki I-451
Tu, Zhuowen II-432
Tucholka, Alan II-117
Tuncali, Kemal I-837

Ugurbil, Kamil II-423
Uhl, Andreas I-247
Unal, Gozde II-776
Unser, Michael II-625
Uzunbaş, M. Gökhan II-34

Vandemeulebroucke, Jef II-365
van der Graaf, Maaike I-869
van der Lugt, Aad I-853
van Geuns, Robert Jan I-369
van Ginneken, Bram I-263, I-747, II-724
Van Leemput, Koen I-272, I-565, II-1075
van Mameren, Henk I-524
van Rikxoort, Eva M. I-263
van Santbrink, Henk I-524
van Vliet, Lucas I-869
van Walsum, Theo I-369
Varadarajan, Balakrishnan I-426
Vasilyev, Nikolay V. I-26
Vass, Melissa I-239
Vavylonis, Dimitrios II-673
Vécsei, Andreas I-247
Vemuri, Baba C. I-190, I-640, I-648
Vercauteren, Tom I-565
Verma, Ragini I-624, I-680
Vernooij, Meike W. I-853
Vetter, Thomas II-1
Vidal, René I-877
Viergever, Max A. I-263, II-724
Villa-Uriol, Maria-Cruz I-140
Vincent, Thomas I-975
Visentini-Scarzanella, Marco I-353
Vitanovski, Dime I-17
Vogel, Jakob II-389
Voigt, Ingmar I-214, II-767
von Elverfeldt, Dominik II-665
Vos, Frans I-869
Vos, Pieter C. II-836

Wachinger, Christian I-779
Wacker, Matthias I-590
Waldman, Stephen D. I-75
Wandell, Brian A. I-861
Wang, Fei I-648
Wang, Fei II-18
Wang, Hanzi I-91
Wang, Lejing I-173
Wang, Linwei II-732
Wang, Peng II-877
Wang, Qian I-656
Wang, Yalin I-337, II-133
Wang, Yang II-767
Warfield, Simon K. I-632, II-300, II-993
Washko, George R. II-690
Wedlake, Chris I-361, I-516
Wei, Yiyi I-198
Weidert, Simon I-173
Wein, Wolfgang I-9, I-763
Wells III, William M. II-109, I-272, II-951
Wels, Michael II-575
Wen, Xu II-339
Werner, Rene I-755
Westin, Carl-Fredrik I-345, II-415, II-690, I-919
Whitaker, Ross I-305, II-167
Wighton, Paul II-1108
Wildenauer, Horst II-860
Wiles, Andrew D. I-361, I-516
Wimmer, Andreas II-26
Wittek, Adam II-300
Wittenberg, Thomas I-459
Wolf, Matthias II-1009
Wollstein, Gadi I-100
Wong, Ken C.L. II-732
Wong, Stephen T.C. II-609
Wong, Tien Yin II-803
Woo, John H. II-141
Woolrich, Mark W. II-514
Wörn, Heinz I-402
Wörz, Stefan I-607
Wrba, Friedrich I-247

Wright, Graham II-750
Wright, Margaret J. I-845, I-967, II-498, II-506
Wu, Binbin II-100
Wu, Guorong I-656, I-721
Wu, Xiaodong II-827

Xia, James II-968
Xu, Chenyang I-9, II-68, II-381
Xu, Guofan II-786
Xu, Xie George II-348
Xuan, Jianhua II-715
Xue, Hui II-741, II-910

Yacoub, Essa II-423
Yamanaka, Noriaki I-83
Yamashita, Hiromasa I-83, I-418, I-451
Yang, Guang-Zhong I-353, I-410, I-483, I-499
Yao, Jianhua II-9
Yap, Pew-Thian I-656, I-721
Yatziv, Liron II-381
Yau, Shing-Tung II-133
Yazdanpanah, Azadeh II-649
Yedidya, Tamir II-976
Yeo, B.T. Thomas I-565, I-598, II-1075
Yezzi, Anthony II-1042
Yi, Zhao II-558
Yin, Youbing I-739
Yip, Michael C. I-26
Ynnerman, Anders I-1000
Yo, Ting-Shuo I-886
Yokota, Futoshi II-811
Yotter, Rachel Aine II-125
Yue, Ning I-43
Yuen, Shelten G. I-26
Yushkevich, Paul A. II-141, II-200, II-466
Yvinec, Mariette II-283

Zachow, Stefan II-76
Zacur, Ernesto II-240
Zecca, Massimiliano I-443
Zellars, Richard I-507
Zhan, Liang I-845
Zhan, Yiqiang II-1033
Zhang, Chong I-140
Zhang, Heye II-732
Zhang, Hui II-141, II-466
Zhang, Tuo II-184
Zhang, Wei I-116
Zhao, Hong II-609
Zhao, Lu II-216
Zhao, Qun I-313
Zheng, Guoyan I-820
Zheng, Yefeng II-575
Zheng, Yuanjie II-852
Zhou, Jinghao I-43
Zhou, Kevin Shaohua I-116, I-255
Zhou, Xiang Sean II-1033
Zhou, Xiaobo II-609, II-968
Zhou, Yan II-1092
Zhu, Hongtu I-721, II-192, II-232
Zhu, Siwei I-845
Zhu, Ying II-877
Zhu, Yun I-206
Zikic, Darko I-828
Zinser, Max II-76
Zuehlsdorff, Sven II-741
Zuo, Siyang I-418