Tianzi Jiang
Nassir Navab
Josien P.W. Pluim
Max A. Viergever (Eds.)

LNCS 6362

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010

13th International Conference
Beijing, China, September 2010
Proceedings, Part II

MICCAI

Springer

Lecture Notes in Computer Science 6362

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

Tianzi Jiang Nassir Navab
Josien P.W. Pluim Max A. Viergever (Eds.)

Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010

13th International Conference
Beijing, China, September 20-24, 2010
Proceedings, Part II

Springer

Volume Editors

Tianzi Jiang
The Chinese Academy of Sciences, Institute of Automation
Beijing 100080, P. R. China
E-mail: jiangtz@nlpr.ia.ac.cn

Nassir Navab
Technische Universität München, Institut für Informatik I16
Boltzmannstr. 3, 85748 Garching, Germany
E-mail: navab@cs.tum.edu

Josien P.W. Pluim
Max A. Viergever
University Medical Center Utrecht, QS.459
Heidelberglaan 100, 3584 CX Utrecht, The Netherlands
E-mail: j.pluim@umcutrecht.nl, max@isi.uu.nl

Library of Congress Control Number: 2010933822

CR Subject Classification (1998): I.4, I.5, I.2.10, I.3.5, J.3, I.6

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

ISSN 0302-9743
ISBN-10 3-642-15744-0 Springer Berlin Heidelberg New York
ISBN-13 978-3-642-15744-8 Springer Berlin Heidelberg New York

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

The 13th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2010, was held in Beijing, China from 20-24 September, 2010. The venue was the China National Convention Center (CNCC), China's largest and newest conference center with excellent facilities and a prime location in the heart of the Olympic Green, adjacent to characteristic constructions like the Bird's Nest (National Stadium) and the Water Cube (National Aquatics Center).

MICCAI is the foremost international scientific event in the field of medical image computing and computer-assisted interventions. The annual conference has a high scientific standard by virtue of the threshold for acceptance, and accordingly MICCAI has built up a track record of attracting leading scientists, engineers and clinicians from a wide range of technical and biomedical disciplines.

This year, we received 786 submissions, well in line with the previous two conferences in New York and London. Three program chairs and a program committee of 31 scientists, all with a recognized standing in the field of the conference, were responsible for the selection of the papers. The review process was set up such that each paper was considered by the three program chairs, two program committee members, and a minimum of three external reviewers. The review process was double-blind, so the reviewers did not know the identity of the authors of the submission.

After a careful evaluation procedure, in which all controversial and gray area papers were discussed individually, we arrived at a total of 251 accepted papers for MICCAI 2010, of which 45 were selected for podium presentation and 206 for poster presentation. The acceptance percentage (32%) was in keeping with that of previous MICCAI conferences. All 251 papers are included in the three MICCAI 2010 LNCS volumes.

We are greatly indebted to the reviewers and to the members of the program committee for their invaluable efforts in critically assessing and evaluating the submissions in a very short time frame.

The annual MICCAI event has, in addition to its main conference, a rising number of satellite tutorials and workshops, organized on the day before and the day after the main conference. This year's call for submission for tutorials and workshops led to a record number of proposals, of which a significant fraction had to be rejected because of space and time limitations. The final program hosted eight tutorials, which together gave a comprehensive overview of many areas of the field, and provided rich educational material especially aimed at PhD students and postdoctoral researchers.

The 15 workshops gave - mostly younger - researchers the opportunity to present their work, often in an early stage of their investigations, so that they could obtain useful feedback from more experienced scientists in the field. The

workshop subjects highlighted topics that were not all fully covered in the main conference, and thus added to the diversity of the MICCAI program. In particular, several workshops offered so-called challenges in which researchers were in competition to best segment or register a set of clinical images with ground truth provided by medical experts. We are grateful to the tutorial and workshop committees, in particular to the chairs Dinggang Shen and Bram van Ginneken, for making these satellite events a success.

Highlights of the conference were the two keynote lectures. Professor Alan C. Evans of the McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada described recent activity in brain network modeling with an emphasis on anatomical correlation analysis in his presentation "Network Analysis of Cortical Anatomy". Professor Guang-Zhong Yang of the Royal Society/Wolfson Medical Image Computing Laboratory, Imperial College, London, UK outlined key clinical challenges and research opportunities in developing minimally invasive surgery systems in his presentation "Snake and Lobster - A Feast for MICCAI?".

MICCAI 2010 would not have been feasible without the efforts of many people behind the scenes. We are particularly indebted to the local organizing committee in Beijing, consisting of Nianming Zuo, Yong Liu, Ming Song, Bing Liu, Bizhen Hong, Shaomei Wang, and Gangqin Zhang, all of the Institute of Automation of the Chinese Academy of Sciences, for their excellent work before and during the conference, and to Jacqueline Wermers for her outstanding assistance with the editorial work in compiling the three Springer LNCS books that contain the proceedings of this conference.

We are obliged to the Board of the MICCAI Society for the opportunity to organize this prestigious conference, and to many of the Society Board and Staff members for valuable and continuous advice and support through all phases of the preparation.

A special word of thanks goes to our sponsors, who generously provided financial support of the conference as a whole, or of specific activities. This greatly helped us with the overall organization of the meeting, as well as allowed us to award prizes for best papers in various categories and travel stipends to an appreciable number of student participants.

It was our great pleasure to welcome the attendees to Beijing for this exciting MICCAI 2010 conference and its satellite tutorials and workshops. The 14th International Conference on Medical Image Computing and Computer-Assisted Intervention will be held in Toronto, Canada, from 15-21 September 2011. We look forward to seeing you there.

September 2010

Tianzi Jiang
Nassir Navab
Josien Pluim
Max Viergever

Organization

General Chair and Co-chairs

Tianzi Jiang	Institute of Automation, CAS, China
Alan Colchester	University of Kent, UK
James Duncan	Yale University, USA

Program Chair and Co-chairs

Max Viergever	Utrecht University and UMC Utrecht, The Netherlands
Nassir Navab	Technische Universität München, Germany
Josien Pluim	University Medical Center Utrecht, The Netherlands

Workshop Chair and Co-chairs

Bram van Ginneken	Radboud University Nijmegen, The Netherlands
Yong Fan	Institute of Automation, CAS, China
Polina Golland	Massachusetts Institute of Technology, USA
Tim Salcudean	University of British Columbia, Canada

Tutorial Chair and Co-chairs

Dinggang Shen	University of North Carolina, USA
Alejandro Frangi	Universitat Pompeu Fabra, Spain
Gábor Székely	ETH Zürich, Switzerland

MICCAI Society, Board of Directors

Nicholas Ayache	INRIA Sophia Antipolis, France
Kevin Cleary	Georgetown University, USA
James Duncan (President)	Yale University, USA
Gabor Fichtinger	Queen's University, Canada
Polina Golland	Massachusetts Institute of Technology, USA
Tianzi Jiang	Institute of Automation, CAS, China
Nassir Navab	Technische Universität München, Germany
Alison Noble	University of Oxford, UK
Sébastien Ourselin	University College London, UK
Ichiro Sakuma	University of Tokyo, Japan
Sandy Wells	Harvard Medical School, USA
Guang-Zhong Yang	Imperial College London, UK

Program Committee

Christian Barillot	IRISA Rennes, France
Albert Chung	Hong Kong UST, China
Gabor Fichtinger	Queen's University, Canada
Alejandro Frangi	Universitat Pompeu Fabra, Spain
Jim Gee	University of Pennsylvania, USA
Bram van Ginneken	Radboud University Nijmegen, The Netherlands
Polina Golland	Massachusetts Institute of Technology, USA
David Hawkes	University College London, UK
Xiaoping Hu	Emory University, USA
Hongen Liao	University of Tokyo, Japan
Huafeng Liu	Zhejiang University, China
Cristian Lorenz	Philips Research Lab Hamburg, Germany
Frederik Maes	University of Leuven, Belgium
Anne Martel	University of Toronto, Canada
Kensaku Mori	Nagoya University, Japan
Mads Nielsen	University of Copenhagen, Denmark
Poul Nielsen	University of Auckland, New Zealand
Wiro Niessen	Erasmus MC Rotterdam, The Netherlands
Xiaochuan Pan	University of Chicago, USA
Franjo Pernuš	University of Ljubljana, Slovenia
Terry Peters	Robarts Research Institute, Canada
Daniel Rueckert	Imperial College London, UK
Tim Salcudean	University of British Columbia, Canada
Yoshinobu Sato	Osaka University, Japan
Dinggang Shen	University of North Carolina, USA
Pengcheng Shi	Rochester Institute of Technology, USA
Gábor Székely	ETH Zürich, Switzerland
Jocelyne Troccaz	TIMC-IMAG, Grenoble, France
Simon Warfield	Harvard University, USA
Carl-Fredrik Westin	Harvard University, USA
Guang-Zhong Yang	Imperial College London, UK

Local Organizing Committee

Nianming Zuo	Institute of Automation,
Yong Liu	Chinese Academy of Sciences, China
Ming Song	
Bing Liu	
Bizhen Hong	
Shaomei Wang	
Gangqin Zhang	
Jacqueline Wermers	UMC Utrecht, The Netherlands

Reviewers

Abolmaesumi, Purang
Abugharbieh, Rafeef
Acar, Burak
Aja-Fernández, Santiago
Akselrod-Ballin, Ayelet
Alexander, Andrew
Alexander, Daniel
Aljabar, Paul
Alomari, Raja
Alvino, Christopher
An, Jungha
Angelini, Elsa
Anor, Tomer
Arbel, Tal
Arridge, Simon
Ashburner, John
Astley, Sue
Atkinson, David
Audette, Michel
Avants, Brian
Awate, Suyash
Babalola, Kolawole
Bach Cuadra, Meritxell
Baillet, Sylvain
Banks, Scott
Barratt, Dean
Batchelor, Philip
Baumann, Michael
Bazin, Pierre-Louis
Beckmann, Christian
Beg, Mirza Faisal
Beichel, Reinhard
Berger, Marie-Odile
Bergtholdt, Martin
Berman, Jeffrey
Betke, Margrit
Bhalerao, Abhir
Bhotika, Rahul
Bian, Junguo
Birkfellner, Wolfgang
Birn, Rasmus
Bischof, Horst
Boctor, Emad
Boisvert, Jonathan
Bosch, Johan
Bouix, Sylvain
Boukerroui, Djamal
Bourgeat, Pierrick
Brady, Mike
Bricault, Ivan
Brun, Caroline
Buelow, Thomas
Bullitt, Elizabeth
Burschka, Darius
Butakoff, Constantine
Cahill, Nathan
Cai, Yiyu
Camara, Oscar
Cardenes, Ruben
Cates, Joshua
Cattin, Philippe
Chakravarty, Mallar
Chen, Elvis
Chen, Sheng
Chen, Wei
Chen, Yunmei
Chen, Zhiqiang
Cheriet, Farida
Chinzei, Kiyoyuki
Chou, Yiyu
Christensen, Gary
Chung, Moo
Cinquin, Philippe
Ciuciu, Philippe
Claridge, Ela
Clarysse, Patrick
Cleary, Kevin
Clerc, Maureen
Colchester, Alan
Collins, Louis
Colliot, Olivier
Comaniciu, Dorin
Commowick, Olivier
Cook, Philip
Cootes, Tim
Cotin, Stéphane
Coulon, Olivier
Coupé, Pierrick

Craddock, Cameron
Crozier, Stuart
Crum, William
Darkner, Sune
Dauguet, Julien
Dawant, Benoit
De Bruijne, Marleen
De Buck, Stijn
De Craene, Mathieu
Deguchi, Daisuke
Dehghan, Ehsan
Deligianni, Fani
Demirci, Stefanie
Deriche, Rachid
Descoteaux, Maxime
Desphande, Gopikrishna
Desvignes, Michel
Dey, Joyoni
Dijkstra, Jouke
DiMaio, Simon
Doignon, Christophe
Douiri, Abdel
Drangova, Maria
Du, Yiping
Duan, Qi
Duchesne, Simon
Duncan, James
Dupont, Pierre
Ebrahimi, Mehran
Ecabert, Olivier
Eggers, Georg
Ehrhardt, Jan
El-Baz, Ayman
Ellis, Randy
Enescu, Monica
Fabry, Thomas
Fahrig, Rebecca
Fan, Yong
Farag, Aly
Fenster, Aaron
Feragen, Aasa
Ferrari, Ricardo
Feuerstein, Marco
Figl, Michael
Fillard, Pierre
Fischer, Bernd
Fitzpatrick, Michael
Fleig, Oliver
Florack, Luc
Fouard, Celine
Freysinger, Wolfgang
Fuernstahl, Philipp
Funka-Lea, Gareth
Gan, Rui
Ganz, Melanie
Gao, Fei
Gee, Andrew
Gerig, Guido
Gessat, Michael
Gholipour, Ali
Gibaud, Bernard
Gladilin, Evgeny
Glocker, Ben
Goksel, Orcun
Gonzalez Ballester, Miguel Angel
Gooding, Mark
Goodlett, Casey
Gooya, Ali
Gorbunova, Vladlena
Grady, Leo
Graham, Jim
Grau, Vicente
Groher, Martin
Gu, Lixu
Guehring, Jens
Guetter, Christoph
Haake, Anne
Hager, Gregory
Hahn, Horst
Hamarneh, Ghassan
Han, Xiao
Hanson, Dennis
Harders, Matthias
Hastreiter, Peter
Hata, Nobuhiko
Haynor, David
He, Yong
Heimann, Tobias
Hellier, Pierre
Heng, Pheng Ann

Hermosillo, Gerardo
Higgins, William
Hipwell, John
Ho, Hon Pong
Hoffmann, Kenneth
Hogeweg, Laurens
Holmes, David
Holz, Dirk
Hoogendoorn, Corné
Hornegger, Joachim
Howe, Robert
Hu, Mingxing
Hu, Zhenghui
Huang, Heng
Huang, Qi-xing
Huang, Xiaolei
Huo, Xiaoming
Hyde, Damon
Ingalhalikar, Madhura
Išgum, Ivana
Jain, Ameet
Janke, Andrew
Jannin, Pierre
Jin, Mingwu
Jomier, Julien
Joshi, Anand
Joshi, Sarang
Kabus, Sven
Kadah, Yasser
Kadir, Timor
Kadoury, Samuel
Kamen, Ali
Kang, Dong-Goo
Karemore, Gopal
Karssemeijer, Nico
Kaus, Michael
Kazanzides, Peter
Keeve, Erwin
Kerrien, Erwan
Kervrann, Charles
Kikinis, Ron
Kim, Boklye
Kindlmann, Gordon
King, Andrew
Kirchberg, Klaus
Kitasaka, Takayuki
Klein, Arno
Klein, Stefan
Klinder, Tobias
Kontos, Despina
Krissian, Karl
Kruggel, Frithjof
Kutter, Oliver
Kybic, Jan
Lai, Shang-Hong
Laine, Andrew
Landman, Bennett
Langs, Georg
Larrabide, Ignacio
Larsen, Rasmus
Lassen, Bianca
Law, Max
Lazar, Mariana
Lee, Junghoon
Leemans, Alexander
Lei, Hao
Lekadir, Karim
Lelieveldt, Boudewijn
Leow, Alex
Lepore, Natasha
Lerch, Jason
Lesage, David
Li, Chunming
Li, Ming
Li, Quanzheng
Li, Shuo
Liang, Jianming
Liao, Rui
Liao, Shu
Likar, Boštjan
Lin, Xiang
Lindseth, Frank
Linguraru, Marius George
Linte, Cristian
Litt, Harold
Liu, Alan
Liu, Tianming
Liu, Yong
Lo, Pechin
Loeckx, Dirk

Loew, Murray
Lu, Le
Luan, Kuan
Luboz, Vincent
Luo, Yishan
Ma, Burton
Madabhushi, Anant
Maeder, Anthony
Magee, Derek
Maier-Hein, Lena
Mainprize, James
Malandain, Gregoire
Manduca, Armando
Mangin, Jean-François
Mao, Hongda
Mao, Hui
Markelj, Primož
Martí, Robert
Martin-Fernandez, Marcos
Masamune, Ken
Masutani, Yoshitaka
Mazza, Edoardo
McClelland, Jamie
McCulloch, Andrew
McGregor, Robert
Metaxas, Dimitris
Metz, Coert
Meyer, Chuck
Miller, James
Milles, Julien
Mohamed, Ashraf
Moireau, Philippe
Mollemans, Wouter
Mungwe, Stanley
Murgasova, Maria
Murphy, Keelin
Mylonas, George
Naish, Michael
Nakamoto, Masahiko
Nash, Martyn
Nedjati-Gilani, Shahrum
Nichols, Thomas
Nicolau, Stephane
Niemeijer, Meindert
Niethammer, Marc
Nimura, Yukitaka
Noble, Alison
Noël, Peter
Nolte, Lutz
Noonan, David
Oda, Masahiro
O'Donnell, Lauren
O'Donnell, Thomas
Ogier, Arnaud
Oguz, Ipek
Olabarriaga, Silvia
Olmos, Salvador
Olszewski, Mark
Orkisz, Maciej
Otake, Yoshito
Ourselin, Sébastien
Ozarslan, Evren
Pang, Wai-Man
Pantazis, Dimitrios
Papadopoulo, Théo
Paragios, Nikos
Pasternak, Ofer
Patriciu, Alexandru
Pavani, Sri Kaushik
Payan, Yohan
Peitgen, Heinz-Otto
Pennec, Xavier
Penney, Graeme
Petersen, Kersten
Petr, Jan
Peyrat, Jean-Marc
Pham, Dzung
Pichon, Eric
Pike, Bruce
Pitiot, Alain
Pizarro, Luis
Pohl, Kilian Maria
Poignet, Philippe
Prager, Richard
Prastawa, Marcel
Prause, Guido
Prima, Sylvain
Prince, Jerry
Promayon, Emmanuel
Qi, Jinyi

Qian, Xiaoning
Radeva, Petia
Rajagopal, Vijayaraghavan
Rajpoot, Nasir
Rangarajan, Anand
Rasche, Volker
Reichl, Tobias
Reinhardt, Joseph
Rexilius, Jan
Reyes, Mauricio
Rhode, Kawal
Ribbens, Annemie
Ridgway, Gerard
Rittscher, Jens
Rivaz, Hassan
Riviere, Cameron
Robb, Richard
Robinson, Emma
Rohlfing, Torsten
Rohling, Robert
Rohr, Karl
Rougon, Nicolas
Rousseau, François
Russakoff, Daniel
Sabuncu, Mert Rory
Sachse, Frank
Sakuma, Ichiro
Salvado, Olivier
Samani, Abbas
Sanchez, Clara
Savadjiev, Peter
Schaap, Michiel
Scherrer, Benoit
Schnabel, Julia
Schweikard, Achim
Sebastian, Rafa
Sermesant, Maxime
Shams, Ramtin
Shechter, Guy
Shi, Yonggang
Shi, Yundi
Shimizu, Akinobu
Siddiqi, Kaleem
Sidky, Emil
Siewerdsen, Jeffrey
Simaan, Nabil
Skrinjar, Oskar
Slagmolen, Pieter
Sled, John
Smal, Ihor
Smeets, Dirk
Smelyanskiy, Mikhail
So, Wai King
Sommer, Stefan
Song, Xubo
Sonka, Milan
Sørensen, Lauge
Spillmann, Jonas
Sporring, Jon
Staal, Joes
Staib, Lawrence
Staring, Marius
Stewart, James
Stoyanov, Danail
Studholme, Colin
Styner, Martin
Suarez, Ralph
Subramanian, Navneeth
Sukno, Federico
Summers, Ronald
Suzuki, Kenji
Szczerba, Dominik
Szilagyi, Laszlo
Tanner, Christine
Tao, Xiaodong
Tasdizen, Tolga
Taylor, Chris
Taylor, Russell
Taylor, Zeike
Tek, Huseyin
Ter Haar Romeny, Bart
Thévenaz, Philippe
Thiran, Jean-Philippe
Thiriet, Marc
Thirion, Bertrand
Todd Pokropek, Andrew
Toews, Matthew
Tomaževič, Dejan
Tosun, Duygu
Tristán-Vega, Antonio

Tsechpenakis, Gavriil
Tustison, Nicholas
Tutar, Ismail
Twining, Carole
Unal, Gozde
Vaillant, Regis
Van Leemput, Koen
Van Rikxoort, Eva
Van Stralen, Marijn
Van Walsum, Theo
Vannier, Michael
Vemuri, Baba
Venkataraman, Archana
Vercauteren, Tom
Verma, Ragini
Vidal, Pierre Paul
Vik, Torbjörn
Vilanova, Anna
Villard, Pierre-Frederic
Von Berg, Jens
Voros, Sandrine
Vos, Frans
Vosburgh, Kirby
Vrooman, Henri
Vrtovec, Tomaž
Wachinger, Christian
Wang, Defeng
Wang, Fei
Wang, Junchen
Wang, Linwei
Wang, Yalin
Wang, Yongmei Michelle
Ward, Aaron
Watton, Paul
Weber, Stefan
Weese, Jürgen
Wein, Wolfgang
Weisenfeld, Neil
Wells, William
West, Jay
Whitaker, Ross
Wiemker, Rafael
Wimmer, Andreas
Wolf, Ivo
Wolz, Robin
Wong, Ken
Woolrich, Mark
Wu, Ed
Wu, Guorong
Wu, John Jue
Xia, Dan
Xu, Jianwu
Xu, Qianyi
Xue, Zhong
Yan, Pingkun
Yang, Hua
Yap, Pew-Thian
Yeo, Thomas
Yezzi, Anthony
Yoo, Terry
Yoshida, Hiro
Young, Alistair
Yu, Weichuan
Yushkevich, Paul
Zang, Yufeng
Zhang, Heye
Zhang, Hui
Zhang, Yong
Zhao, Fuqiang
Zheng, Bo
Zheng, Guoyan
Zheng, Yefeng
Zhou, Luping
Zhou, Kevin
Zhou, Xiang
Zhou, Yu
Zhu, Hongtu
Zhu, Yun
Zikic, Darko
Zöllei, Lilla
Zuo, Nianming
Zwiggelaar, Reyer

Awards Presented at the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI 2009, London

MICCAI Society Enduring Impact Award
The Enduring Impact Award is the highest award of the Medical Image Computing and Computer-Assisted Intervention Society. It is a career award for continued excellence in the MICCAI research field. The 2009 Enduring Impact Award was presented to **Ron Kikinis**, Harvard Medical School, USA.

MICCAI Society Fellowships
MICCAI Fellowships are bestowed annually on a small number of senior members of the Society in recognition of substantial scientific contributions to the MICCAI research field and service to the MICCAI community. The first fellowships were presented in 2009, to
Nicholas Ayache (INRIA Sophia-Antipolis, France)
Alan Colchester (University of Kent, UK)
Takeyoshi Dohi (University of Tokyo, Japan)
Guido Gerig (University of Utah, USA)
David Hawkes (University College London, UK)
Karl Heinz Höhne (University of Hamburg, Germany)
Ron Kikinis (Harvard Medical School, USA)
Terry Peters (Robarts Research Institute, Canada)
Richard Robb (Mayo Clinic, USA)
Chris Taylor (University of Manchester, UK)
Russ Taylor (Johns Hopkins University, USA)
Max Viergever (University Medical Center Utrecht, The Netherlands).

MedIA-MICCAI Prize
The 2009 MedIA-MICCAI Prize for the best paper in the special MICCAI issue of Medical Image Analysis, sponsored by Elsevier, was awarded to
Vicky Wang (University of Auckland, New Zealand)
for the article "Modelling passive diastolic mechanics with quantitative MRI of cardiac structure and function", authored by Vicky Y. Wang, Hoi I. Lam, Daniel B. Ennis, Brett R. Cowan, Alistair A. Young, and Martyn P. Nash.

Best Paper in Navigation
The prize for the best paper in the MICCAI 2009 conference in the area of navigation, sponsored by Medtronic, was awarded to
Wolfgang Wein (Siemens Corporate Research, Princeton, USA)
for the article: "Towards guidance of electrophysiological procedures with real-time 3D intracardiac echocardiography fusion to C-arm CT", authored by Wolfgang Wein, Estelle Camus, Matthias John, Mamadou Diallo, Christophe Duong, Amin Al-Ahmad, Rebecca Fahrig, Ali Khamene, and Chenyang Xu.

Best Paper in Computer-Assisted Intervention Systems and Medical Robotics
The prize for the best paper in the MICCAI 2009 conference in the area of computer-assisted intervention systems and medical robotics, sponsored by Intuitive Surgical, was awarded to
Marcin Balicki (Johns Hopkins University, USA)
for the article "Single fiber optical coherence tomography microsurgical instruments for computer and robot-assisted retinal surgery", authored by Marcin Balicki, Jae-Ho Han, Iulian Iordachita, Peter Gehlbach, James Handa, Russell Taylor, Jin Kang.

MICCAI Young Scientist Awards
The Young Scientist Awards are stimulation prizes awarded to the best first authors of MICCAI contributions in distinct subject areas. The nominees had to be a full-time student at a recognized university at - or within the two years before - the time of submission. The 2009 MICCAI Young Scientist Awards were presented to
Tammy Riklin Raviv (MIT, USA), for the article "Joint segmentation of image ensembles via latent atlases"
Christopher Rohkohl (Friedrich-Alexander University, Germany), for the article "'Interventional 4-D motion estimation and reconstruction of cardiac vasculature without motion"
Peter Savadjiev (Harvard Medical School, USA), for the article "Local white matter geometry indices from diffusion tensor gradients"
Lejing Wang (TU Munich, Germany), for the article "Parallax-free long bone X-ray image stitching"
Yiyi Wei (INRIA Lille, France; LIAMA CASIA, China), for the article "Toward real-time simulation of blood-coil interaction during aneurysm embolization".

Table of Contents – Part II

Ultrasound Imaging

Neuroimage Analysis

Simulation of Anatomical Structures

Endoscopic and Microscopic Imaging

Image Registration

Temporal Diffeomorphic Free-Form Deformation for Strain Quantification in 3D-US Images

Mathieu De Craene[1,2], Gemma Piella[1,2], Nicolas Duchateau[1,2], Etel Silva[4], Adelina Doltra[4], Hang Gao[5], Jan D'hooge[5], Oscar Camara[1,2], Josep Brugada[4], Marta Sitges[4], and Alejandro F. Frangi[1,2,3]

Center for Computational Imaging & Simulation Technologies in Biomedicine (CISTIB)
[1] Information and Communication Technologies Department, Universitat Pompeu Fabra, Barcelona, Spain
[2] Networking Center on Biomedical Research - CIBER-BBN, Barcelona, Spain
[3] Institucio Catalana de Recerca i Estudis Avancats, Barcelona, Spain
[4] Hospital Clínic; IDIBAPS; Universitat de Barcelona, Spain
[5] Department of Cardiovascular Diseases, Cardiovascular Imaging and Dynamics, Katholieke Universiteit Leuven, Belgium

Abstract. This paper presents a new diffeomorphic temporal registration algorithm and its application to motion and strain quantification from a temporal sequence of 3D images. The displacement field is computed by forward eulerian integration of a non-stationary velocity field. The originality of our approach resides in enforcing time consistency by representing the velocity field as a sum of continuous spatiotemporal B-Spline kernels. The accuracy of the developed diffeomorphic technique was first compared to a simple pairwise strategy on synthetic US images with known ground truth motion and with several noise levels, being the proposed algorithm more robust to noise than the pairwise case. Our algorithm was then applied to a database of cardiac 3D+t Ultrasound (US) images of the left ventricle acquired from eight healthy volunteers and three Cardiac Resynchronization Therapy (CRT) patients. On healthy cases, the measured regional strain curves provided uniform strain patterns over all myocardial segments in accordance with clinical literature. On CRT patients, the obtained normalization of the strain pattern after CRT agreed with clinical outcome for the three cases.

1 Introduction

Quantification of cardiac motion and strain provides insight about cardiac function by estimating how a given pathology affects global or local contractility of the myocardium. In clinical routine, motion and strain are usually derived from ultrasound (US) images for which 3D acquisitions are now currently available with sufficient spatiotemporal resolution for characterizing motion and strain. Nonetheless, 3D-US images have lower signal-to-noise ratio (SNR) and temporal resolution than the 2D ones, thus making their processing more challenging.

Several approaches [1,2,3] have been proposed to extend 2D speckle tracking techniques to 3D and to recover myocardial motion from a sequence of 3D-US

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 1–8, 2010.

images. One of the main drawbacks of these techniques based on *pairwise* registrations is that they do not make use of the temporal information embedded in the 3D-US sequences. Parametric spatiotemporal models providing continuous and smooth transformations were proposed by Ledesma-Carbayo *et al.* [4] to exploit temporal information in 2D-US image sequences. However, ensuring temporal smoothness of the displacement field does not render properly *temporal consistency*, i.e. that the displacement field at each time point is related to all the previous times.

Diffeomorphic registration algorithms ([5,6] among others) ensure a continuous, differentiable and with continuous inverse, correspondence between the features to register. Thus, they are particularly well suited to handle medical image sequences as they conserve the topology and the orientation of the observed anatomical structures along time. By integrating a velocity field over time, they provide an elegant way of encoding temporal consistency.

This concept was applied by Khan *et al.* [7] to monitor growth processes by extending the LDDMM [5] image registration algorithm. Velocities were computed using a dense grid, which did not guarantee their spatiotemporal continuity, unlike parametric registration methods. Moreover, a regularization term added to the image similarity metric involved a smoothing kernel that enforced the spatial continuity of the computed velocities. The *temporal* continuity of the velocities was then fully conditioned by the conservation of the topology of the observed features along the image sequence. This assumption does not hold for noisy image sequences, such as cardiac US images. Similar concepts were applied to 2D contours and 3D shapes by Durrleman *et al.* [8]. However, while the computational cost of dense velocity fields is acceptable for sparse topologies, its extension to dense volumetric spatiotemporal data remains critical.

In this paper, we propose a diffeomorphic registration algorithm that models velocities continuously in time and space. We refer to our approach as Temporal Diffeomorphic Free-Form Deformation (TDFFD) algorithm. We extend the popular parametric FFD registration technique [6] by summing spatiotemporal B-Spline kernels to model the 3D+t velocity field. One of the main advantages of our approach is the enforcement of the continuity of the velocity field by using a continuous parametric representation. As a result, the velocity and displacement fields can be computed at any time within the temporal interval captured by the image sequence. The advantage of applying such transformation model for strain quantification purposes is demonstrated here on synthetic and real 3D-US image sequences, with the underlying objective of accurately estimating the impact of Cardiac Resynchronization Therapy (CRT) on 3D strain.

2 Registration Algorithm and Strain Computation

In this paper, we consider a sequence of N images $\{I_n, n = 1 \dots N\}$, each image being defined on a spatial domain $\Omega \subset \mathbb{R}^d$ where $d = 3$ stands for the spatial dimension. Each image I_n is associated to a time instant $t_n \in [0, T]$, with $T > 0 \in \mathbb{R}$. The purpose of the registration algorithm is to solve for the diffeomorphic

mapping $\boldsymbol{\varphi} : \Omega \times [0,T] \rightarrow \mathbb{R}^d$ that relates any point $\mathbf{x}$ in the Eulerian space of coordinates of the first image in the sequence (here taken by convention as $t = 0$) to a *continuous* time $t \in [0,T]$.

The temporal dimension is introduced into the diffeomorphic registration problem by relating the mapping $\boldsymbol{\varphi}$ at any time $t \in [0,T]$ to a time-varying velocity field. In this paper, the velocity field is represented as a sum of spatiotemporal kernels. The B-Spline velocity weights given to all kernels are concatenated in a vector of parameters $\mathbf{p}$, the velocity being then denoted as $\mathbf{v}(\mathbf{x},t,\mathbf{p})$ and computed as

$$\mathbf{v}(\mathbf{x},\mathbf{p},t) = \sum_{\tau}\sum_{\mathbf{c}} \beta\Big(\frac{t-t^{\tau}}{\Delta^{\tau}}\Big) B\Big(\frac{\mathbf{x}-\mathbf{x}^{c}}{\boldsymbol{\Delta}^{\mathbf{c}}}\Big) \mathbf{p}^{\tau,\mathbf{c}} \tag{1}$$

where $B(\cdot)$ is the 3D tensor product of 1D cubic B-spline kernels $\beta(\cdot)$, defined on a sparse grid of 4D control points, being $\mathbf{c}$ the spatial index, τ the temporal index, and $(\boldsymbol{\Delta}^{\mathbf{c}}, \Delta^{\tau})$ the width of the kernels in each dimension. Hence, the $\boldsymbol{\varphi}$ mapping is obtained as

$$\boldsymbol{\varphi}(\mathbf{x},t,\mathbf{p}) = \mathbf{x} + \int_0^t \mathbf{v}(\boldsymbol{\varphi}(\mathbf{x},\tau,\mathbf{p}),\tau,\mathbf{p})\mathrm{d}\tau\,. \tag{2}$$

Forward eulerian integration scheme. The transport equation for computing $\boldsymbol{\varphi}$ in Eq. 2 is solved numerically using a forward Euler integration scheme in which the continuous integral is replaced by a discrete summation. The continuous time interval is now split into a collection of $t_k \in [0,T]$ values where the time increment Δt_k between consecutive time-steps is adapted to ensure invertibility as described in the next subsection. Using this discretization, Eq. 2 can be approximated by

$$\boldsymbol{\varphi}(\mathbf{x},t_n,\mathbf{p}) = \mathbf{x} + \sum_{k=0}^{n-1} \mathbf{v}(\boldsymbol{\varphi}(\mathbf{x},t_k,\mathbf{p}),t_k,\mathbf{p})\Delta t_k\,, \tag{3}$$

If we define $\mathbf{x}_k(\mathbf{p}) \doteq \boldsymbol{\varphi}(\mathbf{x},t_k,\mathbf{p})$, $t_k \doteq \sum_{l=0}^{k-1}\Delta t_l$, and $\mathbf{v}_{k-1}(\mathbf{p}) \doteq \mathbf{v}(\mathbf{x}_{k-1},t_{k-1},\mathbf{p})$ then we can recursively write $\mathbf{x}_k$ as follows:

$$\mathbf{x}_k(\mathbf{p}) = \mathbf{x}_{k-1}(\mathbf{p}) + \mathbf{v}_{k-1}(\mathbf{p})\Delta t_{k-1}\,. \tag{4}$$

Adaptive time-step computation. The integration of Eq. 2 using the discrete approximation of Eq. 4 requires to select a time-step sufficiently small for ensuring accurate computation and invertibility of the mapping $\boldsymbol{\varphi}$. In our method, we start with a uniform sampling of the time interval $[0,T]$, arbitrarily chosen as half of the temporal spacing of the image sequence. To ensure invertibility, one needs to consider the Jacobian of the mapping $\mathbf{x}_n(\mathbf{p})$, here denoted as $\mathbf{D}\mathbf{x}_n(\mathbf{p})$ and computed from Eq. 4 using

$$\mathbf{D}\mathbf{x}_n(\mathbf{p}) = \prod_{k=0}^{n-1}(\mathbf{I} + \mathbf{D}\mathbf{v}_k(\mathbf{p})\Delta t_k) \doteq \prod_{k=0}^{n-1}\Delta\boldsymbol{\varphi}_k\,, \tag{5}$$

where $\mathbf{I}$ stands for the identity matrix. This Jacobian must be positive definite everywhere to ensure invertibility of the transformation. A necessary condition

for this is to have $\det(\mathbf{D}\mathbf{x}_n(\mathbf{p})) > 0$ for all $\mathbf{x} \in \Omega$. Computing the product over k of all $\det(\Delta\boldsymbol{\varphi}_k)$ gives the determinant of the Jacobian matrix in Eq. 5. When a negative value of $\det(\Delta\boldsymbol{\varphi}_k)$ is detected, the value of Δt_k is reduced by a factor 2 until the obtention of a positive determinant.

Similarity metric and non-linear optimization. 3D-US images are characterized by a speckle spatial distribution inside the myocardial wall that is conserved along time. The Mean Squared Error (MSE) appears therefore as a good fit for capturing the optimal set of B-Spline velocity weights $\mathbf{p}$ from Eq. 1. Similarity is measured between the first image in the sequence and all the consecutive frames according to

$$MSE(\mathbf{p}) = \sum_{n=0}^{N-1} \int_{\Omega} \Big(I_0(\mathbf{x}) - I_n(\mathbf{x}_n(\mathbf{p}), t_n) \Big)^2 \mathrm{d}\mathbf{x} \tag{6}$$

Since the number of parameters characterizing the transformation is large, and the metric is explicitly differentiable, gradient-based optimization methods are well indicated for minimizing Eq. 6. In this paper, the L-BFGS-B method was used, which searches the optimum according to the gradient and a low-rank approximation of the Hessian of the metric. For computing the total derivative of Eq. 6 with respect to the weights $\mathbf{p}$ of velocity kernels, we need the following derivative:

$$\frac{\mathrm{d}\mathbf{x}_n}{\mathrm{d}\mathbf{p}} = \frac{\mathrm{d}\mathbf{v}_{n-1}}{\mathrm{d}\mathbf{p}} \Delta t_{n-1} + \frac{\mathrm{d}\mathbf{x}_{n-1}}{\mathrm{d}\mathbf{p}} \text{ where } \frac{\mathrm{d}\mathbf{v}_{n-1}}{\mathrm{d}\mathbf{p}} = \frac{\partial \mathbf{v}_{n-1}}{\partial \mathbf{x}_{n-1}} \frac{\mathrm{d}\mathbf{x}_{n-1}}{\mathrm{d}\mathbf{p}} + \frac{\partial \mathbf{v}_{n-1}}{\partial \mathbf{p}} . \tag{7}$$

Hence, $\mathrm{d}\mathbf{x}_n/\mathrm{d}\mathbf{p}$ can be obtained from the following recursive equation:

$$\frac{\mathrm{d}\mathbf{x}_n}{\mathrm{d}\mathbf{p}} = \frac{\mathrm{d}\mathbf{x}_{n-1}}{\mathrm{d}\mathbf{p}} \Big(\mathbf{I} + \mathbf{D}\mathbf{v}_{n-1} \Delta t_{n-1} \Big) + \frac{\partial \mathbf{v}_{n-1}}{\partial \mathbf{p}} \Delta t_{n-1} , \tag{8}$$

where $\mathbf{D}$ is the Jacobian on all spatial dimensions (i.e., $(\mathbf{D}\mathbf{v})_{ij} = \partial \mathbf{v}_i / \partial \mathbf{x}_j$).

Strain computation. The Cauchy strain tensor is estimated directly using the spatial derivatives of the displacement field from Eq. 5, obtained at any spatiotemporal location using Eq. 4. The Cauchy strain tensor is then computed as

$$\boldsymbol{\sigma}(\mathbf{x}, t_n) = \frac{1}{2} ((\mathbf{D}\mathbf{x}_n)^t \mathbf{D}\mathbf{x}_n - \mathbf{I}) . \tag{9}$$

Strain is obtained along a specific direction $\mathbf{h}$ using $\sigma_{\mathbf{h}}(\mathbf{x}, t) = \mathbf{h}^t \cdot \boldsymbol{\sigma}(\mathbf{x}, t) \cdot \mathbf{h}$. The $\mathbf{h}$ directions considered here are the three vectors (circumferential, longitudinal, radial) of a parabolic coordinate system related to the anatomy of the left ventricle.

3 Experiments

The proposed registration algorithm was applied to three different synthetic datasets to evaluate its accuracy and was then applied to clinical routine 3D-US

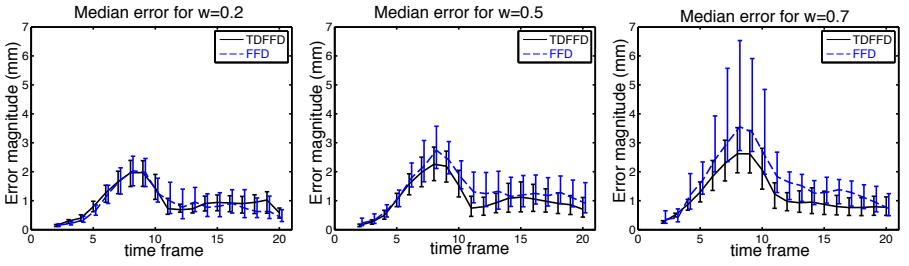


Fig. 1. Median value of error magnitude on the displacement field for the FFD (plotted in blue dashed) and the TDFFD algorithm (black). Vertical bars indicate the second and third quartiles. Three levels of noise are considered: $w = 0.2$ (left), $w = 0.5$ (center) and $w = 0.7$ (right).

images. First, a simple pairwise strategy and the proposed temporally consistent scheme were compared taking a ground truth deformation from simulated US data as reference. Then, strain curves were estimated from 3D-US sequences for 8 volunteers and 3 CRT patients before the therapy and at 12-months follow-up.

3.1 Registration Accuracy on Simulated US Data

Elen *et al.* [2] simulated Left Ventricle (LV) deformation in which the LV was represented as a thick-walled ellipsoid with physiologically relevant end-diastolic dimensions. A simplified kinematic model with an ejection fraction of 60% over a cardiac cycle was used to generate the ground truth displacement field. We used this ground truth data to evaluate the accuracy of the proposed algorithm and compare it to a pairwise registration strategy at different noise levels. Various signal to noise ratios were generated by adjusting intensities inside and outside the myocardial wall using a weight w ($w = 0.2$, $w = 0.5$ and $w = 0.7$ in this paper). Fig. 1 shows the median magnitude and dispersion of the difference between the ground truth displacement field and the ones given by two algorithms: a pairwise FFD (see reference in [6]) and the temporal diffeomorphic FFD registration (TDFFD) algorithms. This error was computed over the entire myocardial wall. FFD pairwise registration was performed between each image and the first image in the sequence, taking the chain of previously computed transformations as bulk transformation. For the two algorithms, the B-Spline grid had an initial resolution of three control points in the longitudinal direction and five in the the two transverse directions. This resolution was then refined twice by a factor 2. For the lowest amount of noise, there was no substantial difference between FFD and TDFFD strategies. However, as the signal to noise ratio decreased, the TDFFD algorithm proved to be more robust and produced smaller errors (maximal median error of 2.6 mm for TDFFD and 3.6 mm for pairwise FFD). The dispersion was also clearly reduced when using the TDFFD algorithm for all noise levels, as observed in Fig. 1. For $w = 0.7$, the upper limit of the third quartile goes from 6.6 mm using pairwise FFD to 3.5 mm using TDFFD.

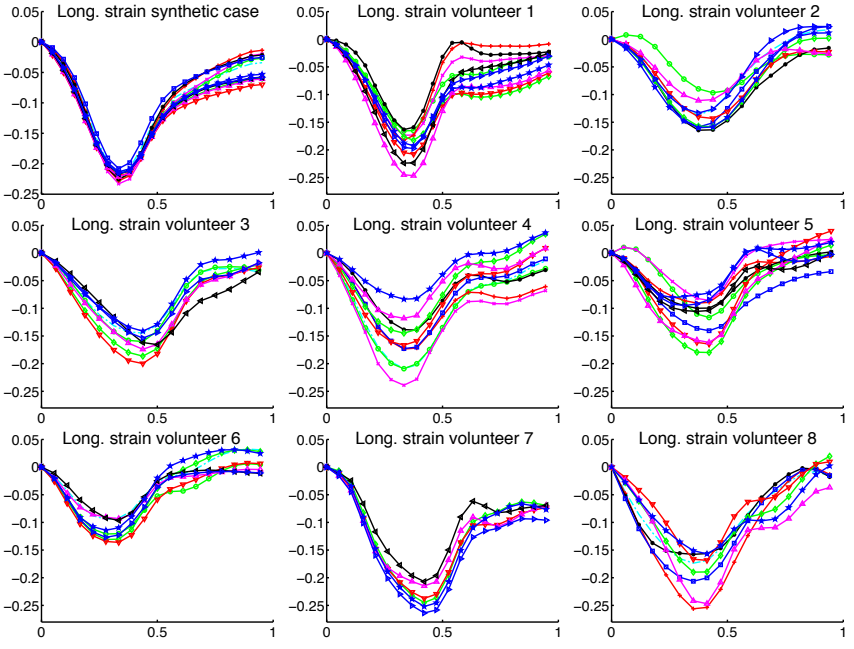


Fig. 2. Longitudinal strain quantified for the synthetic case and 8 healthy volunteers in basal and mid segments. The AHA segments are labelled according to the following: · Basal anterior (1), ∘ Basal anteroseptal (2), × Basal inferoseptal (3), + Basal inferior (4), ∗ Basal inferolateral (5), □ Basal anterolateral (6), ◇ Mid anterior (7), ▽ Mid anteroseptal (8), △ Mid inferoseptal (9), ◁ Mid inferior (10), ▷ Mid inferolateral (11), ⋆ Mid anterolateral (12). Color version of this figure available at http://bit.ly/miccai10.

3.2 Experiments on Clinical Datasets

Data acquisition. We acquired 3D echocardiographic sequences in an apical view for two populations, using a General Electric (Milwaukee, WI, USA) Vivid 7 device. The first population was made up of 8 healthy volunteers (aged 31 ± 6 years), and the second population was composed of 3 CRT patients (aged 61 ± 8 years), who were all clinical responders to CRT. The average number of images per cardiac cycle was of 17.8 for the healthy subjects and 18.3 for the CRT patients. The pixel spacing was on average of $0.9 \times 0.6 \times 0.9\,mm^3$ for the healthy volunteers and $1.0 \times 0.7 \times 1.0\,mm^3$ for the CRT patients.

Strain in healthy volunteers. Fig. 2 shows the recovered longitudinal strain curves for the database of 8 healthy subjects at mid and basal segments of the American Heart Association (AHA). The segments either not totally included in the field of view of the 3D-US images or suffering from typical image artifacts (non visibility of lateral wall, reflections of surrounding anatomical structures, lower spatial resolution on the lateral sides of the sector) of 3D-US acquired

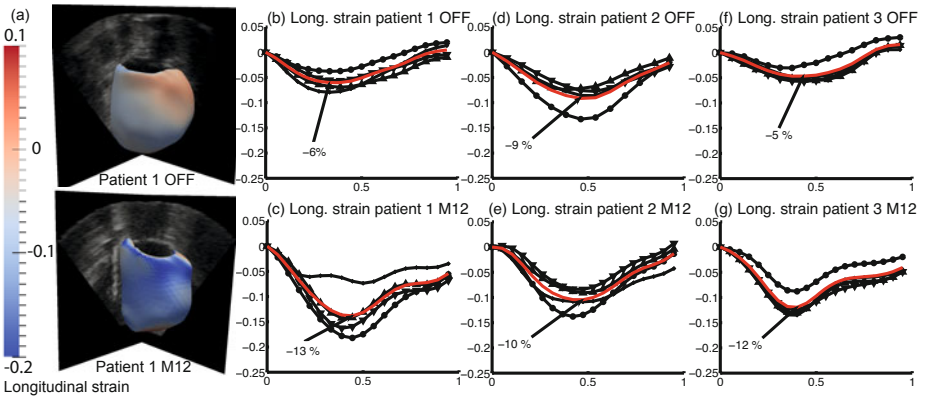


Fig. 3. (a) Longitudinal strain as a colormap at end of systole before CRT (top) and at twelve months follow-up (bottom) for Patient 1. (b-g) Longitudinal strain curves in septal regions before (OFF) and at 12 months follow-up (M12) for Patients 1 to 3. Color version of this figure available at `http://bit.ly/miccai10`.

clinically, were excluded from the analysis. The recovered strain curves showed a similar pattern in all volunteers, in good agreement with clinical literature [9]. The average peak systolic strain was of $-16.3 \pm 4.7\%$. This value is close the $-17.5 \pm 4\%$ reported in [9] and obtained from tagged MRI images. Different phases of diastole such as the isovolumetric relaxation and the atrial contraction (acceleration of the strain at the end of the diastole) periods can be distinguished in cases with higher temporal image resolution such as Volunteers 1 and 4.

Strain before and after CRT. The three CRT patients processed in this paper had dilated geometry before implantation, thus the LV did not fit entirely in the field of view. Therefore, strain was only quantified in the septal regions that usually have the best image quality. Fig. 3 shows longitudinal strain plotted using a color map for the first patient and its temporal evolution per septal segment for all the patients. Average strain curves are shown in red and the peak systolic strain value before and after CRT is indicated by an arrow. Patients 1 and 3 showed a significant improvement in peak systolic strain after CRT that correlated well with an important reverse remodeling observed in these two patients at the 12-month follow-up (reduction of end-systolic volume of 27.7% and 51.0%, respectively). For Patient 2, no substantial change in peak strain value was observed from the strain curves. This patient had lower reverse modeling as observed at the follow-up with a reduction of end-systolic volume of 16.9%. This value is very close to 15 %, which is the threshold used at our institution for defining positive CRT response.

4 Conclusion

This paper presents a new diffeomorphic registration method ensuring temporal consistency of the resulting deformation fields, which can be particularly

useful in image sequences with substantial amount of noise and artifacts. This algorithm was applied for the quantification of strain in 3D US using synthetic datasets, and a set of healthy subjects and CRT patients. Experiments on synthetic US datasets proved an improved robustness and accuracy at high noise levels compared to classical pairwise approaches. On healthy volunteers, the method provided physiologically meaningful longitudinal strain curves with small dispersion among LV segments. On CRT patients, improved peak systolic longitudinal strain in the septum agreed with positive clinical response and reverse remodeling. Future work will include US-adapted similarity metrics and the extension of this registration framework to incorporate compounding strategies to address field of view issues in 3D-US sequences of heart failure patients with dilated LV.

Acknowledgments. This research has been partially funded by the Industrial and Technological Development Center (CDTI) under the CENIT-cvREMOD program and by the European Commission's project euHeart (FP7-ICT-224495). G. Piella and O. Camara were supported by the Ramón y Cajal Program from the Spanish Ministry of Science and Innovation. A. Doltra was supported by a Post-Residency Award from Fundació Clínic.

References

1. Grau, V., Becher, H., Noble, J.: Registration of multiview real-Time 3-D echocardiographic sequences. IEEE Trans. Med. Imag. 26(9), 1154–1165 (2007)
2. Elen, A., Choi, H., Loeckx, D., Gao, H., Claus, P., Suetens, P., Maes, F., D'hooge, J.: Three-dimensional cardiac strain estimation using spatio-temporal elastic registration of ultrasound images: a feasibility study. IEEE Trans. Med. Imag. 27(11), 1580–1591 (2008)
3. Kawagishi, T.: Speckle tracking for assessment of cardiac motion and dyssynchrony. Echocardiography 25(10), 1167–1171 (2008)
4. Ledesma-Carbayo, M., Kybic, J., Desco, M., Santos, A., Suhling, M., Hunziker, P., Unser, M.: Spatio-temporal nonrigid registration for ultrasound cardiac motion estimation. IEEE Trans. Med. Imag. 24(9), 1113–1126 (2005)
5. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. Int. J. Comput. Vision 61(2), 139–157 (2005)
6. Rueckert, D., Aljabar, P., Heckemann, R.A., Hajnal, J., 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)
7. Khan, A., Beg, M.: Representation of time-varying shapes in the large deformation diffeomorphic framework. In: ISBI, pp. 1521–1524. IEEE, Los Alamitos (2008)
8. Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., Ayache, N.: Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 297–304. Springer, Heidelberg (2009)
9. Edvardsen, T., Gerber, B.L., Garot, J., Bluemke, D.A., Lima, J.A., Smiseth, O.A.: Quantitative Assessment of Intrinsic Regional Myocardial Deformation by Doppler Strain Rate Echocardiography in Humans: Validation Against Three-Dimensional Tagged Magnetic Resonance Imaging. Circulation 106(1), 50–56 (2002)

Tracked Ultrasound Elastography (TrUE)

Pezhman Foroughi[1], Hassan Rivaz[1], Ioana N. Fleming[1],
Gregory D. Hager[1], and Emad M. Boctor[1,2]

[1] Dept. of Computer Science, Johns Hopkins University, Baltimore, MD, USA
[2] Dept. of Radiation Oncology, Johns Hopkins University, Baltimore, MD, USA

Abstract. This paper presents a robust framework for freehand ultrasound elastography to cope with uncertainties of freehand palpation using the information from an external tracker. In order to improve the quality of the elasticity images, the proposed method selects a few image pairs such that in each pair the lateral and out-of-plane motions are minimized. It controls the strain rate by choosing the axial motion to be close to a given optimum value. The tracking data also enables fusing multiple strain images that are taken roughly from the same location. This method can be adopted for various trackers and strain estimation algorithms. In this work, we show the results for two tracking systems of electromagnetic (EM) and optical tracker. Using phantom and *ex-vivo* animal experiments, we show that the proposed techniques significantly improve the elasticity images and reduce the dependency to the hand motion of user.

Keywords: Ultrasound, Elastography, Elasticity, Tracking, Strain.

1 Introduction

Ultrasound elastography is an emerging medical imaging modality which involves imaging the mechanical properties of tissue and has numerous clinical applications. Among many variations of ultrasound elastography [1], our work focuses on real-time static elastography, a well-known technique that applies quasi-static compression of tissue and simultaneously images it with ultrasound. Within many techniques proposed for static elastography, we focus on freehand palpation elasticity imaging which involves deforming the tissue by simply pressing the ultrasound probe against it. Freehand ultrasound elastography has shown great potential in clinical applications especially for diagnosis and screening of breast lesions [2]. The application of elastography is not limited to breast, and other applications such as diagnosis of prostate cancer, monitoring ablation and deep vein thrombosis have also been studied.

Despite the reports on success of elastography, yet it has not become a part of any routine clinical application. The main reason is that elastography is highly qualitative and user-dependent. The best result is achieved when the user compresses and decompresses the tissue uniformly in the axial direction with the proper hand motion. It is difficult to control the compression rate as it is governed by the hand motion and the frame rate of RF data. Also, small lateral

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 9–16, 2010.

or out-of-plane motions can compromise the quality of images. However, it is difficult to induce pure axial motion with freehand compression. Sophisticated algorithms can only partially address the problem by compensating for in-plane motions and applying smoothness constraints. The images are also hard to interpret, and artifacts –caused by failure of the strain estimation algorithm or poor hand motion– may be mistaken for lesions inside the soft tissue. Developing an elastography technique that is not affected by poor hand motion and other sources of signal decorrelation will pave the way for wide-spread clinical use of elastography.

To improve the reliability, quality metrics such as persistence in strain images have been developed [3,4]. This quality indicator is calculated for each image and provided to the user as feedback. Persistence is also used to merge multiple elasticity images together [3]. To measure the persistence, strain is computed for two pairs of echo frames, and the resulting images are correlated. Although these techniques offers a major advantage, there remains several limitations. First, the strain has to be estimated before the calculation of the quality metric. With typical ultrasound settings, the frame rate can reach more than 30 Hz. For subsequent frames, an efficient implementation of this image-based metric might cope with this rate. Nonetheless, the task will be extremely difficult to try all the combinations in a series of frames. Moreover, the quality metric will not be able to provide feedback to the user whether he/she should adjust the palpation in certain direction. Also, there would be minimal control over the strain rate.

The ultrasound probe is often tracked in navigation/guidance systems to provide spatial information, to form freehand 3D ultrasound, or to facilitate multimodality registration. In this work, we exploit the tracking data to enhance the quality of the elasticity images. We use the tracking data to select multiple image pairs that contain the optimum deformation for the elastography algorithm. The optimum value for lateral and out-of-plane motions is zero, and the optimum axial motion is determined by the specific elastography algorithm used, which is Normalized Cross-Correlation (NCC) in this work. Next, we fuse the strain images obtained from the multiple image pairs together based on the location of each strain image to improve image quality. We assume that the ultrasound data is 2D. Nonetheless similar techniques proposed here could be extended to 3D ultrasound.

2 Methodology

Consider a sequence of RF data collected during the palpation of tissue using a tracked transducer. We have previously shown that it is possible to synchronize the RF frames with the tracking information relying only on the same data collected during palpation [5]. From synchronization, the tracking information is interpolated at the incident time of each frame. The input to our algorithm is then a series of RF frames along with their corresponding transformation.

First, we need to define a distance function between two frames of RF data. For this purpose, we use a model of image decorrelation in presence of out-of-plane and lateral motion. RF signal is often modeled as the collective response

of scatterers randomly distributed within the resolution cell of the ultrasound [6,7]. Each scatterer is assumed to have an amplitude governed by the shape of the resolution cell and a phase which is distributed from 0 to π uniformly at random. Considering a Gaussian shape for the resolution cell Prager *et. al* [8] calculated the correlation as a function of out-of-plane motion to be $exp(-\frac{\delta^2}{2\sigma^2})$. δ and σ denote the displacement and the width of the resolution cell respectively. Although this function is only valid for fully developed speckle, it provides a convenient estimate of correlation. It should be noted that in [8], the displacement is estimated from correlation, whereas here, we intend to define an energy function based on displacement. Extending this formula to both out-of-plane and lateral displacements, we define our energy function, $E(x, z)$, as follows:

$$E(D_x, D_z) = exp(-K_x \cdot D_x^2 - K_z \cdot D_z^2), \tag{1}$$

where D_x and D_z represent the displacement in out-of-plane and lateral directions. E does not depend on axial motion (D_y) since displacement in axial direction is necessary for strain estimation. K_x and K_z determine the sensitivity to a certain direction. In order to be able to use this function, we need a component-wise metric representing the distance of two frames given their homogeneous transformations. The first step is to compute the relative transformation between them. Suppose $a = [a_x\ a_y\ a_z]^T$ is the axis-angle representation of the relative rotation, and $t = [t_x\, t_y\, t_z]^T$ is the relative translation. Assuming a small rotation, the relative displacement of a point, $P = [x\, y\, 0]^T$, will be $d = a \times P + t$. We then define the distance vector of two frames, $D = [D_x\, D_y\, D_z]^T$, as the RMS of the components of d for all the points in the region of interest (ROI):

$$\begin{aligned}
D_x &= sqrt\{\ \frac{1}{(y_2 - y_1)} \int_{y_1}^{y_2} (-a_z \cdot y + t_x)^2 dy\ \},\\
D_y &= sqrt\{\ \frac{1}{(x_2 - x_1)} \int_{x_1}^{x_2} (a_z \cdot x + t_y)^2 dx\ \},\\
D_z &= sqrt\{\ \frac{1}{(y_2 - y_1)(x_2 - x_1)} \int_{x_1}^{x_2} \int_{y_1}^{y_2} (a_x \cdot y - a_y \cdot x + t_z)^2 dy dx\ \},
\end{aligned} \tag{2}$$

where $sqrt\{.\}$ returns the root. Here, ROI is assumed to be rectangular and determined by x_1, x_2, y_1, and y_2. The vector D provides a measure of distance for each direction separately. We use this vector in Equation (1) which gives us an estimate of "pseudo-correlation" over the ROI.

The data goes through four stages of processing to create a single high-quality strain image. In the first step, few images are selected from the data series that are approximately collected from one cross-section of tissue with minimal lateral and out-of-plane motion. To this end, the energy function of each frame is computed with respect to all other frames in the sequence. Then, the total energy is found for each frame as the sum of the energies of the M closest frames, where closeness implies higher energy, and M is the maximum number of frames to be selected. Then, the frame with the highest total energy (the center frame)

is identified, and the M closest frames to the center frame including itself are selected. Additionally, the frames that have E of less than 0.5 with respect to the center frame are disqualified. This is applied to ensure lower number of frames are chosen when M frames from one cross-section are not available.

In the next stage, the program evaluates all possible combination of frame pairs for elastography. For M frames, there will be $\binom{M}{2} = M(M-1)/2$ pair combinations which will be compared using a slightly modified version of E. Since the pairs are directly compared, it suffices to minimize the exponent of Equation (1) in order to maximize E. We also add a term for axial motion that penalizes compressions that are higher than an optimum compression value, t_{opt}. Hence, a "cost function", $C1$, is defined as follows:

$$C1(D) = K_x \cdot D_x^2 + K_y \cdot \tilde{D_y}^2 + K_z \cdot D_z^2, \quad \tilde{D_y} = \begin{cases} D_y - t_{opt}, & |D_y - t_{opt}| > 0 \\ 0, & |D_y - t_{opt}| \leq 0 \end{cases} \tag{3}$$

where t_{opt} implies the optimal strain, which can be theoretically defined as described in [9]. Here, t_{opt} is set depending on the robustness of the elasticity estimation algorithm. Its value might be within the range of the resolution of the tracker. Therefore, at this stage we do not assign a penalty for the compressions less than t_{opt}. If the compression is close to zero, the contrast of the reconstructed image degrades. The program filters the pairs with low compression in the next stage using image content. Similar to the first part, a maximum number of frames with lowest cost are selected provided that the cost is lower than a threshold. The threshold is not strict to ensure acceptable pairs are not filtered.

The final pairs are selected by recovering the global lateral motion and compression by matching the two RF frames in each pair. The tracking information is used to initialize the search. For instance, the search range for compression is set to be from zero to the tracker reading in axial direction padded in both sides with the maximum error of the tracker. Given two frame I_1 and I_2, the amount of lateral motion a, and compression, b, is found by solving cross-correlation:

$$\arg\max_{a,b} \left\{ \sum_{x,y \in G} I_1(x,y) \cdot I_2(x+a, by) + I_1(x-a, -by) \cdot I_2(x,y) \right\}. \tag{4}$$

The RF data is normalized with standard variation and assumed to have zero mean. We employ two tricks which extensively increases the speed of search. First, we do not match the entire image to solve for these parameters. Instead, only pixels on a grid, G, are used as described by Equation (4). The two terms of Equation (4) ensures that the search remains reciprocal, which means switching the images only affects the sign of a and b. Second, a is recovered by matching only the top part of the two images while b is fixed to one. The reason is that the displacement due to compression is minimal in that region.

Having the global motions, the cost function is modified to penalize very low compressions:

$$C2(\widehat{D}) = K_x \cdot \widehat{D_x}^2 + K_y \cdot \frac{|\widehat{D_y} - t_{opt}|^3}{\widehat{D_y} + c} + K_z \cdot D_z^2, \tag{5}$$

where $\widehat{D_x}$ and $\widehat{D_y}$ are the global motions from Equation (4) converted to mm. c is a small number that limits the cost of zero compression. Finally, the pairs with the lowest cost are selected until a maximum number of frame pairs is reached or the minimum cost grows higher than the average cost.

The last step involves computing the strain for all the selected frame pairs. We have implemented normalized cross-correlation (NCC) [10] to recover the displacements and least squares estimation to calculate the strain. Before calculating strain, the global lateral motion and compression from the previous step are compensated in one image using cubic interpolation. This is known to reduce the error of strain estimation [11]. The final strain image, S_{final} is the weighted average of all the strains:

$$S_{final} = \frac{\sum_{i=1}^{m} w_i \cdot S_i}{\sum_{i=1}^{m} w_i}, \; w_i = \begin{cases} \frac{\rho_i}{1-\rho_i}, & \rho_i > 0.7 \\ 0, & otherwise \end{cases} \tag{6}$$

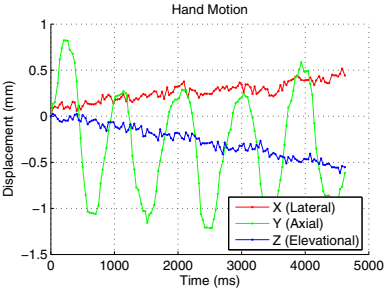
where ρ_i is the correlation coefficient for the ith pair after applying the displacements, and m is the number of pairs. Fusing the strains in this fashion is acceptable since the algorithm only allows for compressions that are close to a predetermined amount optimal for strain estimation.

3 Experiments and Results

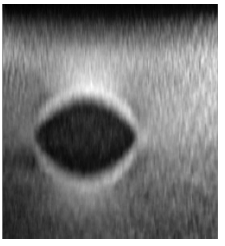
We acquired ultrasound data using a SONOLINE Antares^{TM} ultrasound system (Siemens Medical Solutions USA, Inc.) with a high-frequency ultrasound transducer (VF10-5) at center frequency of 6-8 MHz. We accessed RF through the Axius Direct^{TM} Ultrasound Research Interface provided by Siemens. Our custom data acquisition program was connected to this interface to send the command for capturing RF data. At the same time, the program collected tracking information from either a "Polaris" optical tracker (Northern Digital Inc., Waterloo, Canada) with passive markers or the "medSAFE" EM tracker (Ascension Tech. Corp.).

RF data and tracking information was captured from a breast phantom containing a harder lesion (CIRS elastography phantom, Norfolk, VA) and *ex-vivo* pig liver. Alginate was injected to the liver to mark a part of liver, and then, that area was ablated. The users were asked to palpate the tissue over the hard lesion in the breast phantom and the ablated lesion in the pig liver while data was being collected. Between 100 to 138 RF frames were acquired with the rate of about 30 frames per second.

The first set of data was captured by an experienced user from the breast phantom. Figure 1(a) shows the translation components of hand motion with respect to the first frame. The axial motion is dominant and there is only a gradual drift in the lateral and elevational directions. Figure 1(b) depicts the high-quality strain image resulting from the TrUE algorithm.

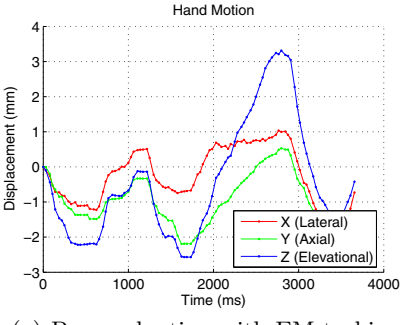


(a) Freehand palpation

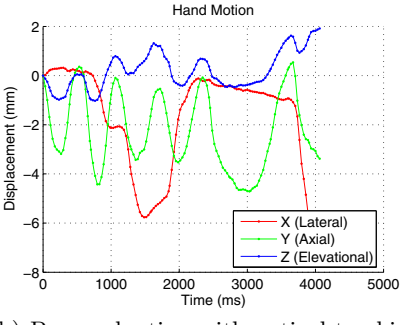


(b) Strain image (TrUE)

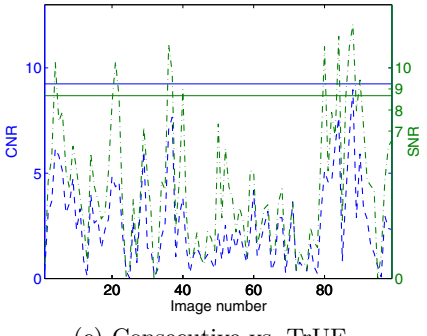
Fig. 1. (a) shows the translation of probe w.r.t. the first image. Proper hand motion is applied as the axial compression is dominant. (b) is the output of our proposed algorithm.



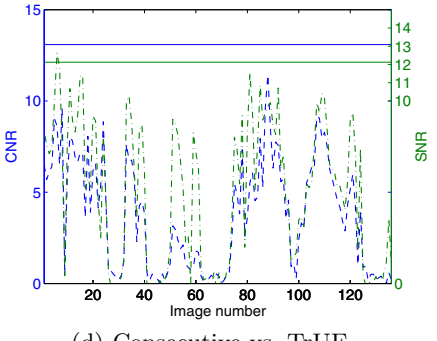
(a) Poor palpation with EM tacking



(b) Poor palpation with optical tracking



(c) Consecutive vs. TrUE



(d) Consecutive vs. TrUE

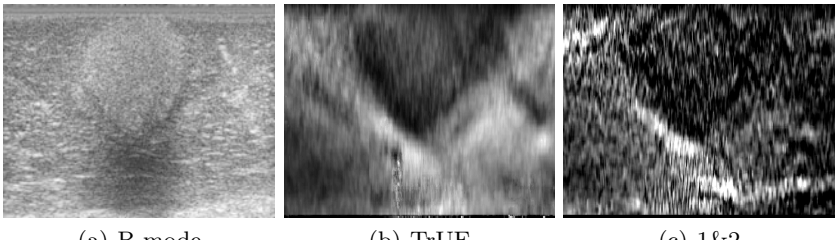
Fig. 2. Two cases of improper motions are shown where the hand motion suffers from large lateral and elevational components evident in relative translations. The results of case 1 with EM tacker is shown on the left column, and the results of case 2 with optical tracker is shown on the right column.

Applying a compression similar to the one shown in Figure 1(a) is a difficult task for novice or even intermediate users. This is especially the case where axial compression does not translate into a simple up and down motion. Ultrasound gel creates a slippery surface that makes the palpation prone to out-of-plane motion. Two case are shown in Figure 2, where one is tracked with the EM tracker and the other one with the optical tracker. In Figure 2(a) the hand motion contains a large amount of out-of-plane motion, whereas, in Figure 2(b), the user has moved the probe laterally. In both cases, the TrUE algorithm generates reliable results. Figures 2 (c) and (d) show the contrast-to-noise ratio (CNR) and signal-to-noise ratio (SNR) of the strain image. The CNR and SNR value are computed from:

$$\text{CNR} = \sqrt{\frac{2(\bar{s}_b - \bar{s}_t)^2}{\sigma_b^2 + \sigma_t^2}}, \quad \text{SNR} = \frac{\bar{s}}{\sigma}, \tag{7}$$

where $\bar{s}$ and σ denote the mean and standard deviation of intensities. The $_t$ or $_b$ subscripts show that the computation is only for the target or the background region, respectively. The SNR and CNR for computing the strain from consecutive frames (the dashed curve) is compared to the SNR and CNR of the strain image from the proposed method (solid line). Using consecutive frames is the standard method of elastography in ultrasound machines. Almost in all cases the TrUE algorithm outperforms the consecutive frames by a large margin.

Although the SNR and CNR provide quantitative measures to compare the strain images, they do not directly reflect the visual quality of strain. In Figure 3, we show results of elastography using our frame selection technique as well as four other strain images calculated from consecutive frames. The Figure shows the effects of improper compression in consecutive frames in the strain image. At the same time our algorithm provides a single reliable strain.



(a) B-mode (b) TrUE (c) 1&2

(d) 2&3 (e) 3&4 (f) 4&5

Fig. 3. Comparison of the strain from TrUE vs. consecutive frames for *ex-vivo* pig liver

4 Discussion

We presented a method of ultrasound elastography which is robust to the quality of the hand motion of the user. Using the information from an external tracker, it automatically selects multiple frame pairs with a specific compression and minimal undesired motions. Our approach does not take into account the tissue motion from other sources such as breathing or patient motion. However, these types of motions are not normally problematic since they occur with a slower pace compared to hand motion.

Our experiments shows that even when the transducer has severe lateral or out-of-plane motions, the algorithm still manages to produce good results. The multi-stage frame selection and careful image fusion makes the TrUE method less sensitive to tacker accuracy and robust to strain estimation failures.

We are planning to use the proposed method in a breast cancer study. For this purpose, we will be implementing our MATLAB code in C. The strain estimation which is still the bottleneck of our approach will be executed in GPU allowing for the use of sophisticated algorithms.

Acknowledgments. Pezhman Foroughi, Hassan Rivaz, and Ioana Fleming are supported by the U.S. Department of Defense pre-doctoral fellowship program.

References

1. Ophir, J., Alam, S., Garra, B., Kallel, F., Konofagou, E., Krouskop, T., Varghese, T.: Elastography: ultrasonic estimation and imaging of the elastic properties of tissues. Annu. Rev. Biomed. Eng. 213, 203–233 (1999)
2. Garra, B., et al.: Elastography of breast lesions: initial clinical results. Radiology 202, 79–86 (1997)
3. Lindop, J.E., Treece, G.M., Gee, A.H., Prager, R.W.: An intelligent interface for freehand strain imaging. Ultrasound Med. Biol. 34, 1117–1128 (2008)
4. Jiang, J., Hall, T.J., Sommer, A.M.: A novel strain formation algorithm for ultrasonic strain imaging. In: IEEE Ultrasonics Symposium, pp. 1282–1285 (2006)
5. Foroughi, P., Hager, G., Boctor, E.: Robust elasticity imaging using external tracker. In: IEEE Int. Symp. Biomed. Imag., pp. 209–212 (2009)
6. Wagner, R., Smith, S., Sandrik, J., Lopez, H.: Statistics of Speckle in Ultrasound B-Scans. IEEE Trans. Sonics and Ultrasonics 17(3), 251–268 (1983)
7. Shankar, P.: A general statistical model for ultrasonic backscattering from tissues. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 47(3), 727–736 (2000)
8. Prager, R., et al.: Sensorless freehand 3-d ultrasound using regression of the echo intensity. Ultrasound Med. Biol. 29, 437–446 (2003)
9. Varghese, T., Ophir, J.: A theoretical framework for performance characterization of elastography: the strain filter. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 44, 164–172 (1997)
10. Céspedes, I., Huang, Y., Ophir, J., Spratt, S.: Methods for estimation of subsample time delays of digitized echo signals. Ultrasound Imaging 17(2), 142–171 (1995)
11. Varghese, T., Ophir, J.: Performance optimization in elastography: Multicompression with temporal stretching. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 18(22), 193–214 (1996)

Evaluation of Inter-session 3D-TRUS to 3D-TRUS Image Registration for Repeat Prostate Biopsies

Vaishali V. Karnik[1,2], Aaron Fenster[1,2,3], Jeff Bax[1,2], Lori Gardi[2], Igor Gyacskov[2], Jacques Montreuil[2], Cesare Romagnoli[2], and Aaron D. Ward[2]

[1] Biomedical Engineering Graduate Program
[2] Robarts Research Institute
[3] Department of Medical Biophysics, The University of Western Ontario, London, ON, Canada, N6A 5C1

Abstract. To ensure accurate targeting and repeatability, 3D TRUS-guided biopsies require registration to determine coordinate transformations to (1) incorporate pre-procedure biopsy plans and (2) compensate for *inter-session* prostate motion and deformation between repeat biopsy sessions. We evaluated prostate surface- and image-based 3D-to-3D TRUS registration by measuring the TRE of manually marked, corresponding, intrinsic fiducials in the whole gland and peripheral zone, and also evaluated the error anisotropy. The image-based rigid and non-rigid methods yielded the best results with mean TREs of 2.26 mm and 1.96 mm, respectively. These results compare favorably with a clinical need for an error of less than 2.5 mm.

Keywords: surface-based registration, image-based registration, accuracy, prostate, repeat biopsy, inter-session, 3D TRUS, peripheral zone.

1 Introduction

Current diagnostic tests for prostate cancer (PCa) include prostate-specific antigen (PSA) tests and digital rectal exams (DRE), but two-dimensional (2D) transrectal ultrasound (TRUS) guided biopsy (Bx) is the clinical standard for definitive diagnosis. 2D TRUS guidance, however, is not without limitations. A 2D view restricts the anatomic information available to the physician for the accurate characterization of tumor location, which is vital for targeting and needle guidance. PCa is particularly difficult to target with 2D TRUS-guided Bx since this cancer tends to be small and multifocal. Furthermore, up to 80% of the cancer can be located in the prostate's peripheral zone (PZ) [1], an area subject to deformation due to ultrasound transducer pressure during the exam (Fig. 1). These difficulties contribute to false negative rates as high as 34% [2], and the detection rate of cancer on the necessary repeat Bx ranges between 10% and 25% [3]. In order to reduce the need for repeat Bxs, and provide better planning for required repeat Bxs, improved techniques are necessary to accurately identify tumor locations and record biopsy cores so as to avoid negative areas and target precancerous lesions. Techniques for obtaining an accurate 3D record of Bx locations and providing guidance of repeated Bxs to specific locations are of

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 17–25, 2010.

profound importance to the timely diagnosis of prostate cancer, and lead to high impact on patient care and quality of life.

A three-dimensional (3D) TRUS-guided prostate biopsy system has been developed in our laboratory, using a mechanically stabilized and tracked end-firing probe that provides a 3D view of the anatomy and records the 3D location of each Bx core (Fig. 1(c)) [4]. In order to accurately transform the 3D locations of Bx cores taken during a previous session into the coordinate space of the current session, 3D-TRUS to 3D-TRUS image registration needs to be performed. The transformation given by this registration will compensate for differences in prostate position and deformation between sessions, which can be substantial due to the time between biopsy sessions, and may be due to differences in patient positioning, edema and swelling due to needle insertion, and bladder and/or rectum filling.

In this study, we evaluate the accuracy of four algorithms for 3D-TRUS to 3D-TRUS registration of prostate images taken during different biopsy sessions. We evaluate rigid and non-rigid variations of both surface-based and image-based registration algorithms. We determine a mean target registration error (TRE) for each algorithm and provide an analysis of the spatial anisotropy in the TRE. Furthermore, since the PZ of the prostate is of particular clinical interest, and we hypothesize that this region may undergo characteristically different (from the remainder of the gland) deformations due to its proximity to the ultrasound transducer, we perform a separate accuracy analysis on this region. Our primary success criterion is a clinically desired TRE of 2.5 mm or less, with a non-dominant fiducial localization error (FLE) [5].

Previous related work focuses on *intra-session* registration of 3D-TRUS to 3D-TRUS [5, 6], and 3D-TRUS to magnetic resonance (MR) [7] prostate images, and accuracies as high as 1.50 mm have been reported. However, to the best of our knowledge, an evaluation of 3D-TRUS to 3D-TRUS *inter-session* registration, with its attendant challenges as described above, has not been previously conducted for image-guided prostate biopsies.

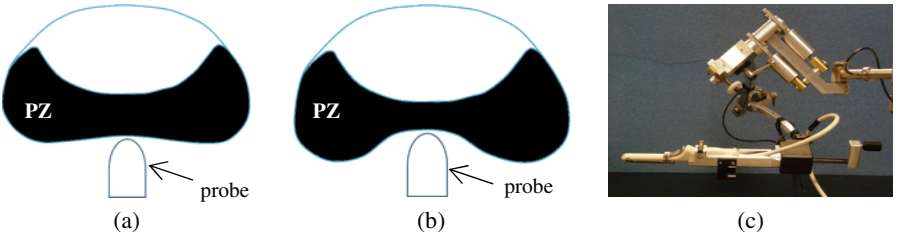


Fig. 1. Axial view of prostate with PZ highlighted in black (a) without deformation and (b) with deformation caused by transducer probe pressure. (c) Mechanically assisted 3D-TRUS biopsy system with real-time tracking and recording of the 3D position of the biopsy needle.

2 Materials and Methods

2.1 Materials

3D TRUS prostate images were obtained from 10 different patients who had undergone two prostate imaging sessions one week apart. The physician inserts the

transducer probe into the patient's rectum and rotates it (by rolling it about the probe axis) to acquire 180 2D images in 1 degree increments. The 2D US images from the US machine are digitized using a frame grabber and reconstructed into a 3D image. Acquisition of the 3D image requires approximately 10 seconds and the reconstruction occurs as the images are acquired, allowing the physician to view the 3D image as it is being acquired. The first set of scans was acquired using a HDI-3500 ultrasound machine and second was acquired using a HDI-5000. The same transducer probe (Philips C9-5) was used with both machines, but calibrated differently. From each patient, 2 images from the two different procedures were selected to be the source (I_S) and target images (I_T). The image dimensions were 448 × 448 × 350 (voxel size of 0.19 × 0.19 × 0.19 mm^3). The selection of images for this study was based on our ability to localize in 3D a set of corresponding intrinsic fiducials (small calcifications) in the image pairs for use in computing the TRE. In all 10 image pairs, we found a total of 92 fiducial pairs, of which 18 were located in the PZ. It is important to note that the fiducials in our study are used only to evaluate the registration methods, and are not used as inputs to the registration algorithms.

2.2 Surface-Based Registration

For surface-based registration, we first segmented the images semi-automatically using a dynamically deformable contour model [8]. We then rigidly registered the surfaces using the iterative closest point (ICP) algorithm [9], where both surfaces are represented by point clouds. In order to align the surfaces, a rigid transformation T is required that maps each point in the source image of patient *i,* (I_S^i), to its corresponding point in the target image of patient *i*, (I_T^i). The algorithm proceeds by (1) establishing correspondence between closest point pairs, and (2) by estimating the transformation required to align the corresponding points. Steps (1) and (2) are iterated until a convergence criterion (RMS of the sum of squares of the closest point distances) of 0.01 mm is met [9].

Following rigid registration, we performed a non-rigid registration using a thin-plate spline (TPS) [10]. Parametric correspondence between surfaces I_S^i and I_T^i was established by radially projecting equal-angle rays from the centroid of the rigidly aligned surfaces, and corresponding the point pair lying on each ray. This approach to correspondence establishment is justified due to the approximately spherical shape of the prostate. A 3D TPS is defined in the 3D image space and aligns the corresponding points. The TPS is applied to I_S^i in order to register it to I_T^i .

2.3 Image-Based Registration

We adopt the block matching approach proposed by Ourselin et al. for rigid registration [11]. The registration consists of the following four steps: (1) dividing the source image into rectangular block regions; (2) for each block in I_S^i , finding the block in a local neighborhood within I_T^i having the optimal value of a chosen image-to-image metric; (3) constructing a vector field mapping each I_S^i block to its best matching I_T^i

block; and (4) performing a least trimmed squares optimization to regularize the vector fields [12] in order to determine the transformation mapping I_S^i to I_T^i. Steps (1) through (4) are iterated, with a progressively smaller block size, in order to provide a coarse-to-fine registration of the images. The chosen image-to-image metric is mutual information [13], due primarily to the effects of directionally-dependent shadowing artifacts present in 3D TRUS images. We used the implementation provided by MedINRIA (INRIA, Asclepios Research Team, 2009) that adopts a multi-resolution approach. Results were achieved using a multi-resolution image pyramid where the finest and coarsest image resolutions used were tuned empirically to 32×32×32 and 16×16×16, respectively. The block size, N, is initialized to $N = \min\left(\frac{X}{8}, \frac{Y}{8}, \frac{Z}{8}\right)$, where *X*, *Y* and *Z* are the 3D image dimensions and the initial search area, Ω, is equal to $2N$ [14]. The parameters are halved when refining the scale and the process stops when the block size is $N = 4$.

For image-based non-rigid registration, the moving image deformation is defined using a regular 3D grid of B-spline control points. We used a limited-memory Broyden Fletcher Goldfarb Shannon optimizer [15], and mutual information [13] as the image-to-image metric. We used the implementation provided by the 3D-SLICER software (Surgical Planning Lab, Harvard Medical School, Boston, USA), which permits the tuning of the following parameters: the number of optimization iterations (we set this parameter to 50), the number of elements along each side of the grid of B-spline control points (10), the number of histogram bins (20) and randomly-sampled voxels (100,000) to use in computing the mutual information metric, and the maximum allowed deformation of the B-spline grid control points (1).

2.4 TRE

We measured the TRE as the overall misalignment of manually marked, corresponding, intrinsic fiducials in I_S^i and I_T^i. The TRE is defined, for the *j*th corresponding fiducial in patient *i*, as the Euclidean distance between the transformed fiducial $\tilde{\boldsymbol{f}}_S^{ij} = T(\boldsymbol{f}_S^{ij})$ from I_S^i and the corresponding fiducial $\boldsymbol{f}_T^{ij}$ in I_T^i, written as $TRE^{ij} = \left\| \boldsymbol{v}^{\mathrm{ij}} \right\|_2$, where $\boldsymbol{v}^{\mathrm{ij}} = (\boldsymbol{f}_T^{ij} - \tilde{\boldsymbol{f}}_S^{ij})$. The mean TRE for patient *i* is denoted as $TRE^i = \sqrt{\sum_j (TRE^{ij})^2 / N_i}$, where N_i is the number of fiducials identified in I_S^i. The overall mean TRE is denoted as $TRE = \sqrt{\sum_i \sum_j (TRE^{ij})^2 / N}$, where N is the total number of fiducial pairs used in the calculation. In this study, there were 92 fiducial pairs in total, of which 18 were within the PZ of the prostates.

2.5 Error Anisotropy

We performed a 3D principal component analysis (PCA) of the 3D point set $\boldsymbol{\Delta} = \left\{ \left(\boldsymbol{f}_{Tx}^{ij} - \tilde{\boldsymbol{f}}_{Sx}^{ij}, \boldsymbol{f}_{Ty}^{ij} - \tilde{\boldsymbol{f}}_{Sy}^{ij}, \boldsymbol{f}_{Tz}^{ij} - \tilde{\boldsymbol{f}}_{Sz}^{ij} \right), \forall i, \forall j \right\}$ in order to evaluate the directional

anisotropy of the TRE. Using MATLAB (The Mathworks Inc., Natick, MA), we computed the PCA by finding the eigenvectors ($\mathbf{e}_i$) and eigenvalues (λ_i) of the covariance matrix of $\mathbf{\Delta}$. This was done for both the whole gland (WG) and PZ. The 95% standard error ellipsoid volumes were defined with their semi-principal axes parallel to the eigenvectors, with lengths $a_i = \sqrt{\lambda_i \times \chi^2_{\alpha,\upsilon}}$, ($i$ = 1, 2, 3 and $\chi_{\alpha,\upsilon} \approx 2.80$), where $\chi^2_{\alpha,\upsilon}$ is the probability distribution evaluated for 95% confidence (α = 0.05) and 3 degrees of freedom (υ = 3) [16].

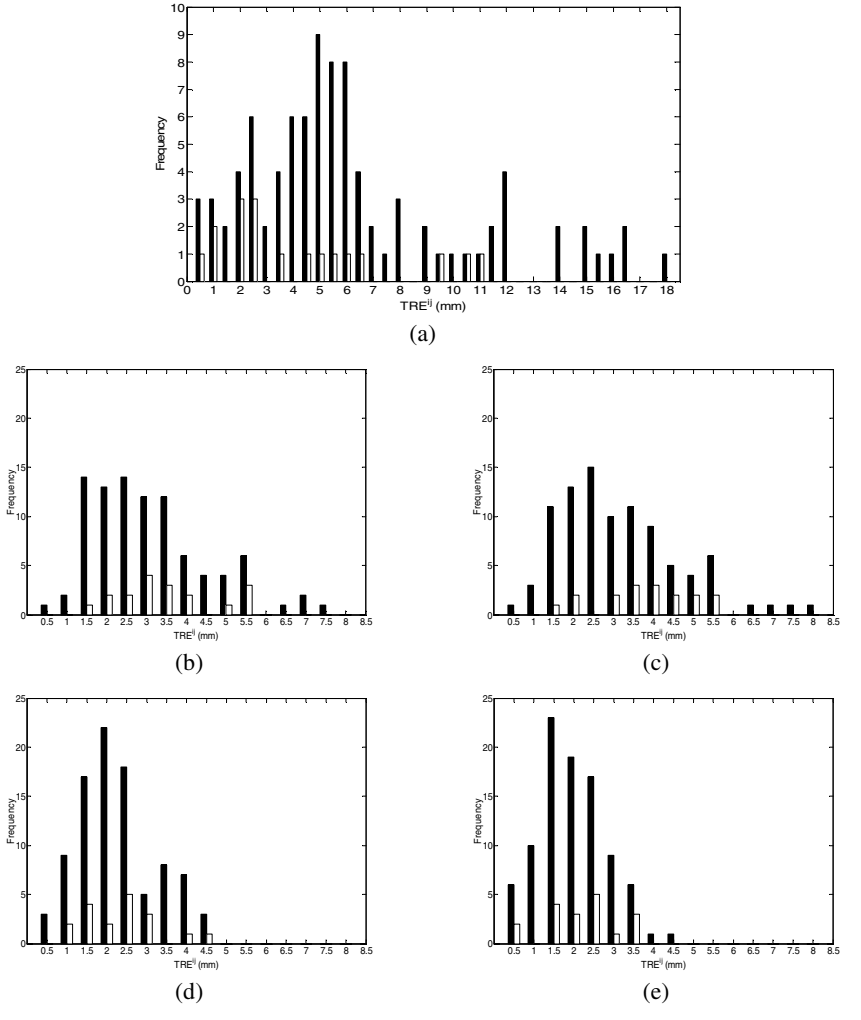


Fig. 2. WG (black) and PZ (white) frequency distributions of **(a)** pre-registered distances between all 92 fiducial pairs, **(b)** rigid surface-based TRE^{ij}, **(c)** non-rigid surface-based TRE^{ij}, **(d)** rigid image-based TRE^{ij}, **(e)** non-rigid image-based TRE^{ij}

3 Results

3.1 TRE

Our pre-registration mean misalignment was ~7.45 mm, ranging from 0.28-18.0 mm. The frequency distributions of the WG and PZ TREs are shown in Figure 2 and the TRE results are summarized in Table 1. The two-tailed paired t-test results shown in Table 2 indicate that surface- and image-based rigid registration TREs were statistically significantly different with $p = 4.29 \times 10^{-6}$. The surface- and image-based non-rigid TREs were also significantly different, with $p = 1.60 \times 10^{-9}$.

Table 1. Whole gland (WG) and peripheral zone (PZ) mean TRE results for surface-based (SB) and image-based (IB) registration

Registration	WG		PZ	
	SB	IB	SB	IB
Pre-registration mean TRE (mm)	7.36 ± 4.17		5.23 ± 3.29	
Rigid mean TRE (mm)	3.20 ± 1.46	2.26 ± 0.98	3.45 ± 1.21	2.30 ± 0.90
Non-rigid mean TRE(mm)	3.29 ± 1.50	1.96 ± 0.85	3.62 ± 1.12	2.11 ± 0.91

Table 2. t-test results for surface-based (SB), image-based (IB), rigid (R), and non-rigid (NR) registration

	SB vs. IB	SB vs. IB	SB	IB
	R	NR	R vs. NR	R vs. NR
WG p-value	4.29×10^{-6}	1.6×10^{-9}	0.05	1.38×10^{-7}
PZ p-value	1.88×10^{-4}	6.21×10^{-6}	0.12	0.07

3.2 Error Anisotropy

Figure 3 shows the 95% confidence ellipsoids for the PCA performed on the directional components of the WG and PZ TREs for the non-rigid registration methods. The WG pre-registered ellipsoid volume (EV) is 4.46 cm^3 while the post-registration EVs range from 0.12 cm^3 to 0.41 cm^3. The EVs for surface-based rigid and non-rigid TRE were greater than the image-based EVs by an average of 0.27 cm^3.

The PZ pre-registered EV is 1.58 cm^3 while the post-registration volumes range from 0.14 cm^3 to 0.37 cm^3. The EVs for surface-based rigid and non-rigid TRE were greater than the EVs for image-based TRE by an average of 0.22 cm^3. Table 3 shows the ratios of the eigenvalues for each registration approach as well as for the unregistered case, to illustrate any anisotropy in the error.

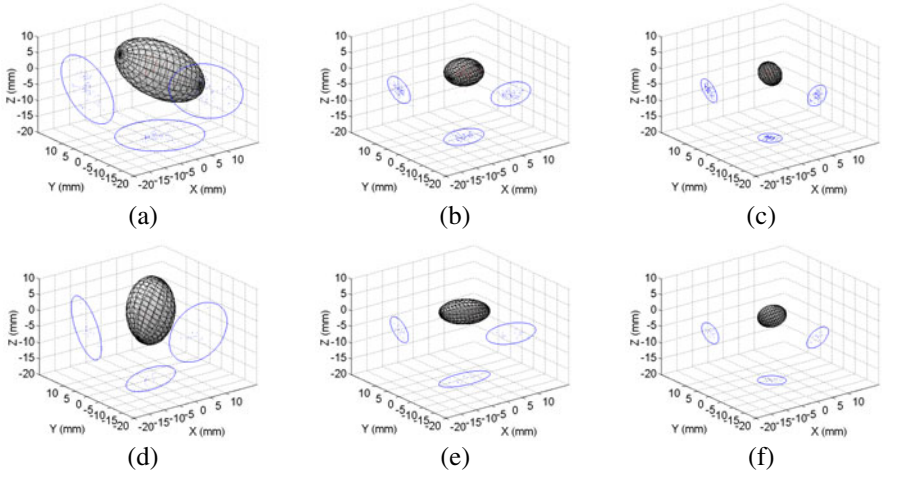


Fig. 3. Error ellipsoids for WG and PZ: **(a)** WG pre-registered data, **(b)** WG surface-based non-rigid registration, **(c)** WG image-based non-rigid registration, **(d)** PZ pre-registered data, **(e)** PZ surface-based non-rigid registration, and **(f)** PZ image-based non-rigid registration

Table 3. Ratios of 95% TRE confidence ellipsoid eigenvalues

	WG			PZ		
Registration Method	λ_1 / λ_2	λ_2 / λ_3	λ_1 / λ_3	λ_1 / λ_2	λ_2 / λ_3	λ_1 / λ_3
Unregistered	1.16	1.37	1.43	1.24	1.72	1.95
Surface-based rigid	1.51	1.36	1.42	2.08	1.6	1.6
Surface-based non-rigid	1.76	1.86	2.02	2.58	2.75	3.13
Image-based rigid	1.16	1.37	1.43	1.24	1.72	1.95
Image-based non-rigid	1.51	1.36	1.42	2.08	1.6	1.6

4 Discussion

Using 3D image registration to correct for prostate motion and deformation that occurs between Bx procedures is vital to accurate Bx planning and needle guidance. The desired TRE is 2.5 mm since the smallest tumor considered clinically significant has a radius of ~5 mm (for a 0.5cm^3 spherical tumor) [17]. Since the TRE (RMSE) provides an estimate of the standard deviation of the normal distribution of biopsy targets given by a registration algorithm, a TRE of 2.5 mm gives a confidence interval with 95.4% of the registered targets within the clinically significant 5 mm radius.

The non-rigid image-based method yielded the lowest WG and PZ TRE values of 1.96 mm and 2.11 mm, respectively. Figures 2(d) and (e) show that the image-based method outperformed the surface-based method by illustrating the right-skewed distribution of the TRE values (for both WG and PZ), with ~65% of the values below the respective means. In a previous evaluation of image registration methods in the *intra-session* context, the non-rigid image-based TRE for images taken 6 minutes apart was

1.50 mm [5]. This shows that after a time gap of 1 week, the same image-based registration algorithm, tuned as in [5], is able to achieve comparable results.

The t-test results in Table 2 reveal statistically significant differences in all cases except when comparing surface-based rigid and non-rigid registration values. We speculate that this could be due to the variability introduced by segmentation (±1.31 mm) [5]. The FLE [18] was 0.21 mm [5], and thus was not dominant.

Figure 3 shows the anisotropy in the TRE where both WG and PZ pre-registered fiducial misalignments are predominantly in the z-direction, which is along the transducer probe axis where we might expect misalignments due to probe pressure. The surface-based errors are predominantly anisotropic in x and y, which may be the result of inconsistent segmentation due to poor image contrast at the prostate boundaries parallel to the ultrasound beam. Table 3 indicates increased error anisotropy in the PZ, in comparison to the WG. This may be due to deformation that characteristically pushes the central region of the PZ upward in response to probe pressure, while the lateral "lobes" of the PZ bulge downward and laterally. The post-registration 3D EVs are dramatically reduced (by a factor of 10–37 for WG, and 4.27–11.29 for PZ) as compared with the pre-registered EVs.

Our results demonstrate that in the *inter-session* scenario with 1 week between sessions, the tested surface-based and image-based 3D TRUS image registration techniques yield clinically sufficient accuracy, with image-based registration giving the lowest error. Furthermore, due to the segmentation step required for surface-based registration, the two methods are comparable in execution time on a contemporary single-core workstation (90–105 s). The image-based method is preferred since it is more accurate, does not require segmentation and can potentially be automated.

References

1. Prostate Cancer Treatment Guide (2008-2010), http://www.prostate-cancer.com
2. Fink, K., et al.: Evaluation of transition zone and lateral sextant biopsies for prostate cancer detection after initial sextant biopsy. Urology 61, 748–753 (2003)
3. Kronz, D., et al.: Predicting cancer following a diagnosis of high-grade prostatic intraepithelial neoplasia on needle biopsy: Data on men with more than one follow-up biopsy. The American Journal of Surgical Pathology 25(8), 1079–1085 (2001)
4. Bax, J., et al.: Mechanically assisted 3D ultrasound guided prostate biopsy system. Medical Physics 35(12), 5397–5410 (2008)
5. Karnik, V.V., et al.: Assessment of image registration accuracy in three-dimensional transrectal ultrasound guided prostate biopsy. Medical Physics 37(2), 802–813 (2010)
6. Baumann, M., et al.: Prostate biopsy assistance system with gland deformation estimation for enhanced precision. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 67–74. Springer, Heidelberg (2009)
7. Hu, Y., et al.: MR to ultrasound image registration for guiding prostate biopsy and interventions. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 787–794. Springer, Heidelberg (2009)
8. Wang, Y., et al.: Semiautomatic three-dimensional segmentation of the prostate using two-dimensional ultrasound images. Medical Physics 30(5), 887–897 (2003)

9. Besl, P., McKay, N.: A method for registration of 3-D shapes. IEEE Trans. on Pattern Analysis and Machine Intelligence 14, 239–256 (1992)
10. Bookstein, F., et al.: Principal warps: Thin-plate splines and the decomposition of deformations. IEEE Trans. on Pattern Analysis and Machine Intelligence 11(6), 567–585 (1989)
11. Ourselin, S., et al.: Block matching: A general framework to improve robustness of rigid registration of medical images. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 557–566. Springer, Heidelberg (2001)
12. Rousseuw, P.J., Leroy, A.M.: Robust regression and outlier detection, 1st edn. Wiley Series in Probability and Mathematical Statistics (1987)
13. Wells III, M., et al.: Multi-modal volume registration by maximization of mutual information. Medical Image Analysis 1(1), 35–51 (1996)
14. Ourselin, S., et al.: Reconstructing a 3D structure from serial histological sections. Image and Vision Computing 19, 25–31 (2000)
15. Zhu, C., et al.: L-BFGS-B: Algorithm 778: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization. ACM TMS 23(4), 550–560 (1997)
16. Irwin, M., et al.: Registered 3-D ultrasound and digital stereotactic mammography for breast biopsy guidance. IEEE TMI 27(3), 391–398 (2008)
17. Epstein, J.I., et al.: Utility of saturation biopsy to predict insignificant cancer at radical prostatectomy. Urology 66(2), 356–360 (2005)
18. Fitzpatrick, J.M., et al.: Predicting error in rigid-body point-based registration. IEEE TMI 17(5), 694–702 (1998)

Manifold Learning for Image-Based Breathing Gating with Application to 4D Ultrasound*

Christian Wachinger, Mehmet Yigitsoy, and Nassir Navab

Computer Aided Medical Procedures (CAMP), TUM, Munich, Germany
{wachinge,yigitsoy,navab}@cs.tum.edu

Abstract. Breathing motion leads to a significant displacement and deformation of organs in the abdominal region. This makes the detection of the breathing phase for numerous applications necessary. We propose a new, purely image-based respiratory gating method for ultrasound. Further, we use this technique to provide a solution for breathing affected 4D ultrasound acquisitions with a wobbler probe. We achieve the gating with Laplacian eigenmaps, a manifold learning technique, to determine the low-dimensional manifold embedded in the high-dimensional image space. Since Laplacian eigenmaps assign each ultrasound frame a coordinate in low-dimensional space by respecting the neighborhood relationship, they are well suited for analyzing the breathing cycle. For the 4D application, we perform the manifold learning for each angle, and consecutively, align all the local curves and perform a curve fitting to achieve a globally consistent breathing signal. We performed the image-based gating on several 2D and 3D ultrasound datasets over time, and quantified its very good performance by comparing it to measurements from an external gating system.

1 Introduction

Imaging organs in thorax and abdomen is affected by respiratory motion. For consecutive processing steps, it is often necessary to assign to each image its corresponding breathing phase. This is achieved with external gating devices, which the patient gets connected to. These devices, however, have long setup times, prolong the overall acquisition, are costly, and consequently, rarely used in practice. Moreover, the synchronization of image data and breathing signal is not trivial. While certain imaging devices, such as CT and MR scanners, support the connection of respiratory gating systems, we are not aware of such a possibility for ultrasound; leaving the synchronization to the user. We propose an image-based respiratory gating system for ultrasound using manifold learning. Moreover, we use this technique to provide a solution for acquiring breathing affected 4D ultrasound with a wobbler probe. The proposed method is fully automatic, and does not need any prior information or training data.

* This work was partly funded by the European Commission. We want to thank Athanasios Karamalis and Oliver Kutter for discussions.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 26–33, 2010.

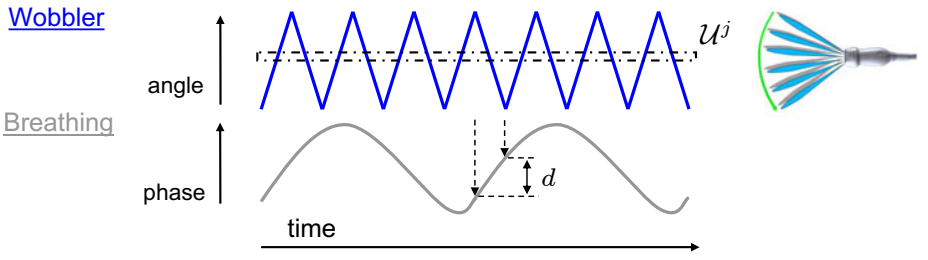


Fig. 1. Wobbler angle (blue) and respiratory phase (gray) over time. Dashed lines indicate respiratory change d within one sweep. Dash dotted line indicates frames from same angle over several breathing cycles.

One of the applications for which the assignment of the respiratory phase is important is 2D and 3D ultrasound mosaicing. In order to achieve good results, images from the same respiratory phase have to be combined. An alternative to breathing gating are breath-hold acquisitions, but they further complicate the procedure and are dependent on the patients ability for breath-hold. Another application that we investigate in more details throughout the article, and for which we have not yet found a solution proposed in the literature, is the acquisition of breathing affected 4D ultrasound with a mechanically steered transducer, also referred to as wobbler. The problems for using a wobbler in such a scenario is that images in one sweep do not contain consistent information, but represent the anatomy in different breathing phases. We illustrate this in figure 1, where we schematically plot the deviation angle of the wobbler together with the respiratory signal over time. The phase difference d indicates the range of breathing phases accumulated in one sweep. We propose to select all frames acquired from the same angle (dash dotted line) and apply the image-based gating on each of those sets of images. Having the respiratory signal estimated for each angle, we align these curves and apply a robust spline curve fitting to create a globally consistent respiratory signal. This, consequently, allows us to reconstruct volumes for specific breathing stages. An alternative to the application of a wobbler would be a native 3D transducer with elements arranged on a 2D array. Those systems, however, are still expensive and the access to their data streaming and radio frequency data is very restricted.

1.1 Related Work

There are several papers on image-based gating in ultrasound for detecting the cardiac motion [1,2,3,4]. These approaches apply techniques that are either (i) specific to detecting the cardiac signal *e.g.* centroid algorithm [4], (ii) based on user interaction [1], or (iii) designed for intravascular ultrasound [2,3]. We are only aware of the work of Sundar *et al.* [5], where an automatic technique for breathing gating in ultrasound is proposed. It bases on the phase correlation technique to estimate the motion between successive frames. The breathing phase is estimated from the energy change between consecutive frames. The inherent

limitation of the phase correlation algorithm is that it finds the *global translation* in the image plane. Considering the case of 2D ultrasound, the organ motion is not necessarily in-plane, and consequently, there is no uniform global translation.

Several manifold learning techniques were proposed in the literature with common techniques being Isomap [6] and Laplacian eigenmaps [7]. Since its introduction, manifold learning has been applied for a multitude of applications, including segmentation [8], registration [9], tracking [10], recognition [11], and 4D CT reconstruction [12]. The approach taken in the reconstruction is similar to ours because Isomap is used to estimate the breathing phase on CT slabs. In our work, we deal with the specific challenges of the integration of 4D ultrasound wobbler data. To this end, we focus on Laplacian eigenamps, since we achieved better results in comparison to Isomap.

2 Method

The general idea of manifold learning is to project a manifold in high dimensional space $\mathbb{R}^N$ to a low dimensional space $\mathbb{R}^n$, while preserving the local neighborhood. For our application, we consider one dimension of the ambient space for each image pixel, so N is corresponding to the resolution of the ultrasound images. For the low dimensional space, we set $n = 1$, because we want to use the coordinate of the points directly as breathing phase estimation. Considering k ultrasound images $\mathcal{U} = \{\mathbf{u}_1, \ldots, \mathbf{u}_k\}$ that are acquired over several breathing cycles, the manifold learning m assigns each image a coordinate in the low dimensional space ϕ_i

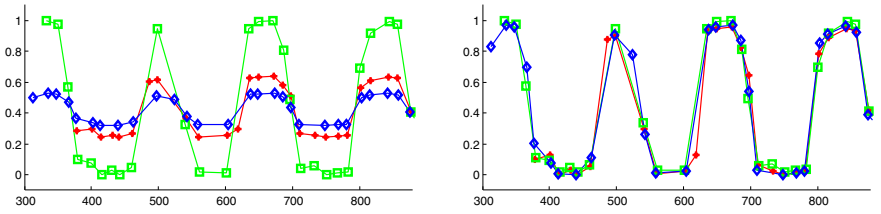
$$m : \mathbb{R}^N \rightarrow \mathbb{R}^1 \tag{1}$$

$$\mathbf{u}_i \mapsto \phi_i, \tag{2}$$

with $1 \leq i \leq k$. The suggestion that ultrasound images lie on a low dimensional manifold in the ambient space seems to be justified because variations between neighboring slices are smooth, and further, slices from the same respiratory phase but different acquisition times are similar. Moreover, since manifold learning techniques try to optimally preserve local information [7], meaning that similar images are mapped to similar positions in the low dimensional space, it is reasonable to use ϕ_i as an estimate for the respiratory phase.

2.1 Laplacian Eigenmaps

We propose the application of Laplacian eigenmaps [7] for the respiratory phase estimation because the technique is well founded on mathematical concepts (Laplace Beltrami operator) and computationally efficient. Laplacian eigenmaps build upon the construction of a graph, which represents the neighborhood information of the data set. Subsequently, the graph Laplacian is applied to calculate a low-dimensional representation of the data that preserves the local neighborhood information in an optimal way.



(a) Not Aligned (b) Aligned

Fig. 2. Breathing signals from manifold learning. Illustrated is the case for 3 angles (3 colors). X-axis indicates ultrasound frame number. The plots show the signals before and after alignment.

We construct a graph with a node for each point $\mathbf{u}_i$ and with edges connecting neighboring nodes. We select for each image $\mathbf{u}_i$ the l nearest neighbors, by evaluating the norm $||\mathbf{u}_i - \mathbf{u}_\iota||^2$. Further, heat kernel-based weights are assigned to the edges with $w_{i\iota} = e^{-||\mathbf{u}_i - \mathbf{u}_\iota||^2/(2\cdot\sigma^2)}$ and σ^2 the variance [7]. The similarity measure is important for neighborhood selection and weighting, where the calculation of the Euclidean norm between the points is equivalent to calculating the sum of squared differences (SSD) between the images. A vast number of similarity measures is proposed in the context of medical image registration. Since we deal with monomodal data for our application, we investigate the performance of SSD and the correlation coefficient (CC). The calculation of CC is similar to the calculation of SSD on normalized input images. Once the neighborhood graph is constructed, the eigenvectors of the graph Laplacian provide the embedding map.

2.2 Global Consistency

After the breathing gating is performed for each angle, we have to establish the correspondence between different angles in order to construct the global respiratory signal. Be $\mathcal{U}$ the set of all acquired images. We partition the set in disjunct subsets $\mathcal{U}^1, \ldots, \mathcal{U}^\alpha$, corresponding to the number of different deflection angles α of the wobbler (dash dotted region in figure 1). We perform the manifold learning for each of the subsets separately $m^j(\mathbf{u}_i) = \phi_i^j$, with $1 \leq j \leq \alpha$. So depending on the acquisition angle of the ultrasound image $\mathbf{u}_i$, the corresponding manifold learning m^j is performed. Considering all the phases estimated from one angle, we have the local respiratory signals $\Phi^j = \{\phi_1^j, \ldots, \phi_v^j\}$, with v the number of frames per angle. Each local signal contains a consistent estimation of the breathing signal. It is, however, not possible to directly compare local signals, because the 1D projection of the manifold learning can be in an arbitrary range. This is illustrated in figure 2(a) with exemplary three local signals corresponding to three angular positions. A simple normalization of each of the local signals Φ^j is not sufficient because the extreme positions of the breathing cycle may not

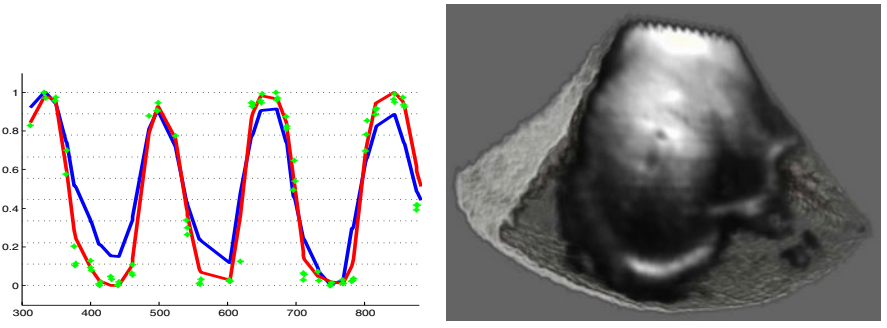


Fig. 3. Aligned local signals (green crosses). Robust spline fitting (red). Ground truth (blue). Dotted lines indicate separation of breathing cycle into several stages. For each stage a volume is compounded, with an exemplary volume rendering on the right image.

be reached within them. Consequently, we affinely register local signals in order to retrieve the best scaling s^j and translation t^j

$$\Phi^j \mapsto s^j \cdot \Phi^j + t^j. \tag{3}$$

This is, in fact, a groupwise registration scenario, where we choose to align each pair of neighboring curves with a pairwise registration, starting from the middle one. The result of the alignment is shown in figure 2(b).

The values of the partial signals Φ^j are now comparable, however, may still contain outliers. Consequently, we apply a robust curve fitting to all the sample points to retrieve the global breathing signal. We experimented with various curve models, including Fourier, sum of sine waves, and splines. We achieved best results with fitting a spline curve because it allows for the most flexibility, which is important due to irregularity of breathing. The value of the fitted curve then represents the breathing phase of the ultrasound frames.

In a final step, the breathing cycle is classified into several breathing stages. For each of the breathing stages, the ultrasound frames along the various angles are gathered, and compounded into a final volume, see figure 3.

3 Experiments

For our experiments we use the ultrasound system from Ultrasonix (Richmond, Canada) and the optical tracking system from A.R.T. (Weilheim, Germany). Both systems are connected to a workstation PC. For the synchronization, we time stamp the data on the tracking system and use a network time server to calculate the time offset. For the ultrasound data, we use the direct streaming of B-mode images over the network. We perform tests on multiple patient datasets acquired from different positions, focusing on the liver and kidney.

In order to validate our results, we compare them to the measurements of an external gating system. In [13], four different gating systems are compared with

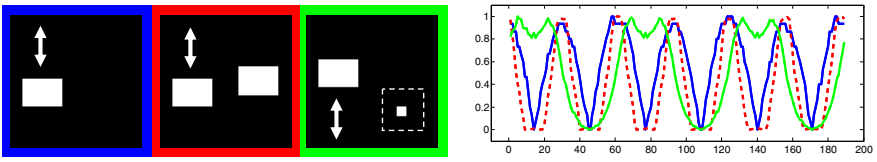


Fig. 4. Analysis of the phase correlation technique for synthetic images. Three different motion scenarios with corresponding energy (gating) curves.

best results for an elastic belt and an optical tracking system. We use the tracking system with markers attached to the chest of the patient. We apply a principal component analysis of the 6D tracking data to find the principal component along whose direction we measure the breathing motion. Further, we low-pass filter the signal to remove cardiac motion and extract the respiratory signal. We refer to the tracked signal as ground truth, which is not completely correct because it contains tracking errors. However, it is the best that can currently be achieved [13] and is sufficient to validate the performance of our image-based approach.

We compare our approach to the phase correlation technique [5][1]. Unfortunately, we do not achieve meaningful results for our datasets. We think that this is due to the limitation of the technique of approximating the 3D motion with a global translation in 2D. In order to illustrate this limitation, we produced synthetic images which show periodic motion. The first scenario consists of a rectangle moving up and down, see figure 4. For the second, we add a fixed rectangle, and for the third we add a rectangle that grows and shrinks (see additional material for videos). We plot the corresponding energy curves of the phase correlation technique. For the first scenario (blue) the signal is correct. The addition of a fixed object (red) already leads to a slight distortion, while the addition of the shrinking/growing object (green), leads to an extraction of a false motion signal. Since already the addition of the shrinking/growing object avoids the extraction of the correct motion, it is comprehensible that this approach is not best suited for breathing estimation in a noisy ultrasound environment with 3D anatomy moving in and out of plane.

The resolution of our ultrasound images is 640×480 pixels. We downsample the images in each direction by a factor of 2, leading to $N = \frac{1}{4} \cdot 640 \cdot 480$. This enables a faster processing and leads to no noticeable degradation of the manifold learning. We show excerpts of two data sets in figure 5. We perform all our experiments with a graph neighborhood of $l = 14$. The number of images for manifold learning varies between 100 and 300, where we did not notice a dependency of the performance on the number of input samples.

In figure 6, we show the result of the respiratory gating for one of the 2D datasets together with the ground truth signal. We also calculate the correlation coefficient for multiple 2D data sets, shown in table 1. It is remarkable that the

[1] We want to thank the authors of [5] for sharing source code.

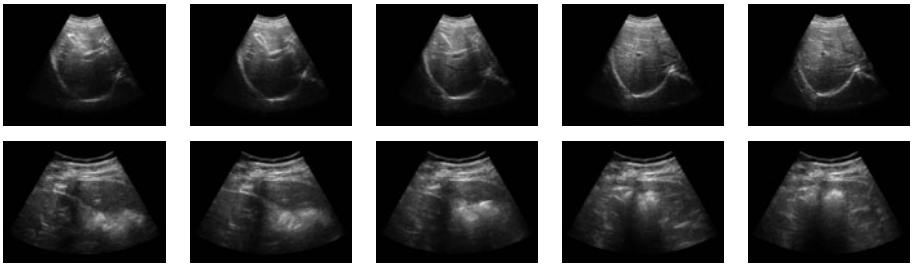


Fig. 5. 2D Ultrasound images over time from liver (abdomen, right upper quadrant, oblique section) and kidney (left lateral decubitus position, right intercostal flank section)

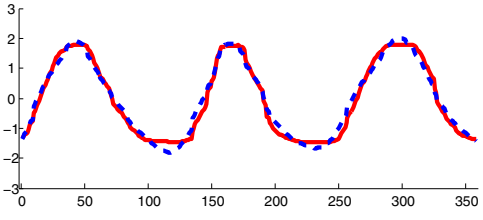


Fig. 6. Breathing gating results for 2D (red: estimated signal, blue: ground truth)

Table 1. 2D

2D Correlation	
liver1	95.4
liver2	94.4
liver3	93.6
kidney	97.3

Table 2. 3D

3D Correlation	
liver 30°	94.3
liver 45°	95.8
liver 60°	96.8
kidney 45°	94.4

ground truth signal is almost perfectly detected. All peaks in the ground truth signal also appear in the detection. Further, the calculation of the correlation, which is in the range of 95%, confirms the visual similarity of the graphs. We also experimented with normalizing the images before passing them to the manifold learning, noticed however no significant improvement.

For the 4D experiments, we show the result of a fitted curve in figure 2. We also calculate the correlation coefficient between the fitted curves and ground truth for four datasets, see table 2. We experimented with three different angular ranges, 30°, 45°, and 60° (maximum of probe), for which the probe steers to 15, 21, and 29 different angular positions. We split the breathing signal into 9 different breathing stages, and compound a 3D volume for each of the stages. A volume rendering of one of the volumes is shown in figure 3. The additional material contains a video showing the 4D volume rendering.

All image-based approaches rely on ultrasound acquisitions from the same position, because otherwise it is not possible to differentiate between probe motion and breathing motion. To investigate this assumption, we attached a tracking target to the transducer and analyzed its trajectory. This analysis showed only a negligible deviation. The still position therefore does not limit the applicability of our method, which is also confirmed by our good gating results.

4 Conclusion

We presented an automatic, image-based respiratory gating method for ultrasound using manifold learning. Moreover, we proposed a solution for acquiring 4D breathing data with a wobbler probe. Our method has the advantage that it is fully automatic and does not require a training phase or prior information about the underlying anatomy. We demonstrated this in our experiments by performing our analysis on various datasets showing different organs and sections. The results of these experiments were very good, for both, 2D and 3D ultrasound data over time. Our approach therefore presents an attractive alternative to external tracking and gating systems with their various setup issues and synchronization problems.

References

1. Treece, G., Prager, R., Gee, A., Cash, C., Berman, L.: Grey-scale gating for freehand 3D ultrasound. In: ISBI, pp. 993–996 (2002)
2. Zhu, H., Oakeson, K., Friedman, M.: Retrieval of cardiac phase from IVUS sequences. In: Proceedings of SPIE, vol. 5035, p. 135 (2003)
3. de Winter, S., Hamers, R., Degertekin, M., Tanabe, K., Lemos, P., Serruys, P., Roelandt, J., Bruining, N.: A novel retrospective gating method for intracoronary ultrasound images based on image properties. In: Computers in Cardiology (2003)
4. Karadayi, K., Hayashi, T., Kim, Y.: Automatic image-based gating for 4d ultrasound. In: Engineering in Medicine and Biology Society (2006)
5. Sundar, H., Khamene, A., Yatziv, L., Xu, C.: Automatic image-based cardiac and respiratory cycle synchronization and gating of image sequences. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 381–388. Springer, Heidelberg (2009)
6. Tenenbaum, J., Silva, V., Langford, J.: A global geometric framework for nonlinear dimensionality reduction. Science 290(5500), 2319 (2000)
7. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. 15(6) (2003)
8. Zhang, Q., Souvenir, R., Pless, R.: On manifold structure of cardiac MRI data: Application to segmentation. In: CVPR, vol. 1, pp. 1092–1098 (2006)
9. Hamm, J., Davatzikos, C., Verma, R.: Efficient large deformation registration via geodesics on a learned manifold of images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 680–687. Springer, Heidelberg (2009)
10. Lee, C., Elgammal, A.: Modeling view and posture manifolds for tracking. In: ICCV (2007)
11. Wachinger, C., Mateus, D., Keil, A., Navab, N.: Manifold Learning for Patient Position Detection in MRI. In: ISBI (April 2010)
12. Georg, M., Souvenir, R., Hope, A., Pless, R.: Manifold learning for 4d ct reconstruction of the lung. In: MMBIA (2008)
13. Martinez-Möller, A., Bundschuh, R., Riedel, M., Navab, N., Ziegler, S., Schwaiger, M., Nekolla, S.: Comparison of respiratory sensors and its compliance for respiratory gating in emission tomography. Journal of Nuclear Medicine (2007)

Measurement of the Skin-Liver Capsule Distance on Ultrasound RF Data for 1D Transient Elastography

Stéphane Audière[1,2], Maurice Charbit[1], Elsa D. Angelini[1], Jennifer Oudry[2], and Laurent Sandrin[2]

[1] Institut Telecom, Telecom ParisTech, CNRS LTCI, Paris, France
[2] Echosens, Research and Development Department, Paris, France

Abstract. Vibration-controlled transient elastography (VCTE™) technique is routinely used in clinical practice to assess non-invasively the liver stiffness which is correlated to hepatic fibrosis. Adequate use of the VCTE™ probe requires the knowledge of the distance between the skin and the liver parenchyma. This paper compares two methods to estimate this distance using spatial variations of the spectral content of ultrasound radiofrequency (RF) lines, obtained from a probe consisting of a single element ultrasound transducer placed in front of the liver right lobe. Results on a database of 188 patients, including normal-weight and obese persons, show that the spectral variance can accurately discriminate the subcutaneous fat from the liver tissue. The proposed algorithm works in real-time and is suitable for VCTE™ scanning protocol setup.

Keywords: ultrasound, RF lines, liver, spectral analysis, elastography.

1 Introduction

Vibration-controlled transient elastography (VCTE™) technique [1] is routinely used in clinical practice to quantify liver stiffness by measuring the velocity of a low-frequency shear wave generated and travelling through the liver. It has been demonstrated that stiffness is highly correlated with fibrosis stage assessed by liver biopsy [2]. The VCTE™ device can be operated with different 1D-ultrasound probes, operating at different frequencies, depending on the patient morphology. The highest frequency probe provides higher spatial resolution but cannot be used on patients with a thick fat layer, due to frequency-dependent ultrasound attenuation in the fat. Therefore, the choice of the probe during an examination depends on the distance between the probe (in contact with the skin surface) and the liver. This distance corresponds to the subcutaneous thickness, called skin-liver capsule distance (SCD). The SCD is currently measured manually on ultrasound B-mode imaging. Automatic measurement of the SCD is therefore of great interest for VCTE™ scanning protocol setup to alleviate manual measurements, reduce operator dependency, and standardize the measurement technique. Ultrasound segmentation methods have mainly focused on B-mode imaging given the difficulty in getting access to raw RF data on commercial systems and the complexity of the RF signal content. Unfortunately, B-mode images are formed from the detection of the RF ultrasound line envelopes, which removes

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 34–41, 2010.

rich spectral information that could be used to characterize soft tissue properties. Among the few works dedicated to RF-based ultrasound segmentation, we can cite Hammoude et al. [3] who used changes in central frequency due to tissue attenuation to segment the heart muscle but had to face erratic changes in ultrasound signal. Boukerroui et al. [4] proposed a segmentation framework based on gray scale levels and texture analysis from the ultrasound envelope signal, combined with tissue characterization from the analysis of the central frequency spectrum of the ultrasound RF lines. More recently, Dydenko et al. [5] combined RF autoregressive spectrum parameters with the estimation of ultrasound propagation speed to characterize cardiac tissues and obtained good results using the estimated speed variance on *in vivo* signals. Moradi et al. [6] used the fractal dimension of RF time series for tissue characterisation, but needed high pulse repetition frequency rates. Davignon et al. [7] used the "integrated backScatter coefficient" (IBS) and the average central frequency to improve a segmentation method based on the ultrasound envelope signal. These studies all confirmed that the spectral information of the ultrasound RF lines was adapted for segmentation tasks, providing discriminant information on the acoustic properties of the different tissues.

To our knowledge, few studies have been performed on the automatic measurement of the thickness of the subcutaneous layer, or on the skin-liver capsule distance. Ng et al. [8] recently showed that changes in the spectral content of 2D-spatial RF images enabled to accurately measure subcutaneous layer thickness on humans, but required spatial compounding. In this paper, a real-time implementation of the spectral analysis proposed by Ng et al. [8], but applied to 1D ultrasound RF lines, was evaluated and compared to the exploitation of the IBS parameter from [7].

2 Materials and Methods

The SCD includes the layers of tissues connecting the epidermis of the skin surface to the liver. Fibrous membranes of connective tissues may vary in thickness and density between individuals. Furthermore, the fibrous membranes may contain varying proportion of fat and are more echogenic than the liver parenchyma, as illustrated in Fig. 1. Overall the subcutaneous layer has a rather heterogeneous appearance while the liver is rather homogeneous.

2.1 Spectral RF Analysis

In most cases, the liver capsule is not precisely detectable on 1D ultrasound RF lines. Indeed, the direction of the ultrasound beam and the surface of the capsule must be perfectly perpendicular to accurately detect the echo generated by the capsule, as illustrated in Fig. 1. Since it is not possible to visually control the orientation of the VCTE™ ultrasound beam with respect to the liver surface, the echo amplitude cannot be directly used to determine the SCD. On the other hand, distinction between the subcutaneous layer and the liver can be performed via spectral analysis of the RF lines, based on the fact that ultrasound attenuation is tissue- and frequency-dependent, as detailed below.

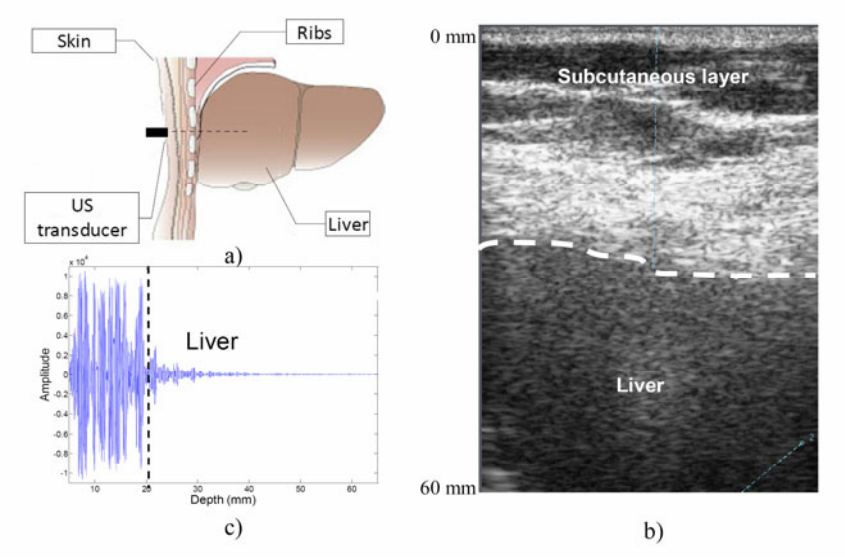


Fig. 1. (a) Positioning of the ultrasound (US) transducer for a VCTE™ and B-mode scanning, at the level of the 7th right intercostal space. (b) B-mode image (the position of the capsule is identified with a dashed line.). (c) Ultrasound RF line acquired with the VCTE™ probe.

The spectrogram $S(f,z)$ of the RF signal at frequency f and at a distance z from the probe is generally modelled as:

$$S(f,z) = |H(f)|^2 |R(f)|^2 |A(f,z)|^2, \quad (1)$$

where H is the transfer function of the ultrasonic source, R is the backscatter component from the echoes and A is the attenuation term. In human soft tissues, ultrasound RF lines are exponentially attenuated in depth, leading to the following model for A:

$$A(f,z) = e^{-\alpha(f)z}, \quad (2)$$

Within the limited bandwidth of the ultrasound transducer, experimental measures have shown that soft tissues create attenuation phenomena of the RF lines which are linearly proportional to its frequency. Huisman et al. [9] have proposed the following attenuation law:

$$\alpha(f) = \alpha_0 + \beta(f - f_c), \quad (3)$$

where α_0 is the tissue-specific attenuation coefficient at the central frequency f_c of the transducer and β is the slope of the frequency-dependent linear attenuation. This attenuation model, exponential in z and with an attenuation coefficient linear in f naturally motivates the choice of a spectral analysis of the ultrasound RF signal for tissue characterization (different tissues being encountered along the z-axis).

RF Spectral parameters: Two parameters were calculated from the spectrograms of acquired RF lines: the total energy reflected by the tissues (measured with the IBS) and the spectral variance. The IBS was proposed in [7] as a suitable discriminant

parameter for tissue characterization. In our case, the tissues constituting the subcutaneous layer are hyper-echogenic relative to the liver parenchyma, leading to potentially higher IBS values. The spectral variance σ^2 characterizes the local variation of the centroid frequency content of the spectrogram. Both can be calculated with the moment method, as proposed in [10], IBS corresponding to m_0 and the spectral variance being computed as:

$$\sigma^2(z) = \frac{m_2(z)}{m_0(z)} - \left(\frac{m_1(z)}{m_0(z)}\right)^2, \tag{4}$$

where

$$m_n(z) = \int_{BW} f^n S(f,z) df, \tag{5}$$

with *BW* being the useful bandwidth of the transducer. The spectral variance σ^2 does not depend on the echogeneity difference between tissues (i. e. RF signal strength), but rather on the heterogeneity of the tissues. From Eq. (3), σ^2 is also independent of α_0 and f_c but only depends on the tissue-dependent attenuation coefficient β.

2.2 Ultrasound RF Scanning Setup and Procedure

The VCTE™ device used in this study was composed of a probe containing a low-frequency vibrator, an ultrasonic transducer operating, depending on the probe, at 2.5 MHz or 3.5 MHz, a dedicated electronic system and a control unit. The sampling frequency of the ultrasound signal was 50 MHz with a 14-bit resolution. A single ultrasound element was used both as an emitter and a receiver. For each volunteer, 400 ultrasound RF lines were acquired, along a fixed scan line, at a pulse repetition frequency of 20 Hz during 20 s.

In this study, the following acquisition protocol was used: the liver was first identified with B-mode ultrasound images and the skin capsule distance was manually measured on the images (as illustrated in Fig. 1). The VCTE™ probe was then placed at the same scanning location, but with the application of a controlled pressure on the skin surface of 20 to 30 kPa, for RF recording in elastography scanning mode. Application of such pressure limits probe motion while the patient breathes.

A database was acquired with the two VCTE™ probes on 188 volunteers: 85% with normal weights and 15% of obese (corresponding to a body mass index (BMI) >30 kg/m²). The ultrasound RF lines were acquired for a depth range between 5 and 65 mm, to accommodate for the variability of the skin-liver capsule distance within the volunteers population.

3 Results and Discussion

3.1 Spectral Analysis Setup

For each RF line, the spectrogram was computed with short-time Fourier transform (STFT). Spatial windows of length 12λ were used for the STFT computation, where λ represents the wavelength (and 2λ the pulse length) of the emitted ultrasound signal

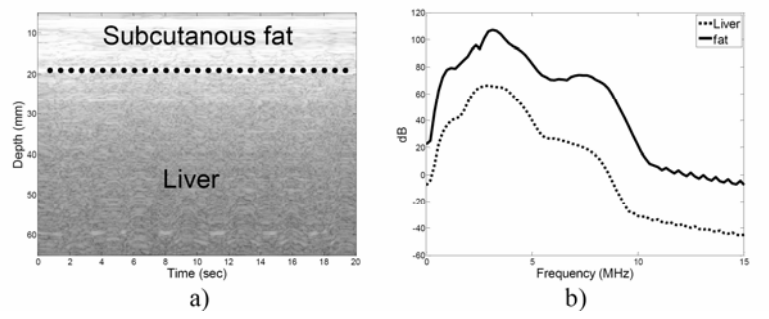


Fig. 2. (a) Series of RF lines acquired with a VCTE™ probe on a moderately obese volunteer. (b) Averaged localized spectrogram profiles inside the liver and inside the subcutaneous layer.

(estimated for an average ultrasound velocity c set to 1540 m/s for human soft tissues). This spatial window size, which sets the axial precision of the spectrogram, was chosen to optimize the tradeoff between frequency and temporal resolution. Overlaps of 75% and a weighting Blackman window were applied. With these parameters, we considered that the spectrogram computation had a precision of 3λ (i.e. 3 times the axial resolution of the US transducer).

In Fig. 2, an example of a series of temporal ultrasound RF lines acquired with a fixed-position VCTE™ 3.5 MHz probe is provided. The volunteer was a male patient measuring 1.67 m and weighing 82 kg (BMI = 30 kg/m², and B-mode SCD = 20 mm, corresponding to moderate obesity). Averaged profiles (over 0.5 s) of localized spectrogram data for the subcutaneous layer (z=10 mm) and inside the liver (z=40 mm) are shown on Fig. 2b. Spectral analysis of the spectrograms, to extract the IBS and spectral variance, was performed on a series of consecutive 400 RF lines acquired over 20 s, maintaining the probe in a fixed position. Spatio-temporal maps of these parameters were computed and a binarization of these maps was applied, based on Otsu's thresholding, to separate the subcutaneous layer from the liver tissue. The series of SCD measures was then combined to provide a robust estimate and compensate for potential tissue interface motion during breathing and to attenuate the effects of noisy components. Averaging of the SCD measures over the series of RF lines was performed in two steps: (1) local averaging over temporal windows of 2 s (approximate breathing period), (2) hysteresis thresholding for outliers removal on the temporal profile of the SCD values and selection of the SCD median value. For real-time computation during VCTE™ scanning, these computations are performed in temporal streaming mode, with SCD measures refreshed every 0.5 s.

Parametric maps of the normalized IBS and the normalized spectral variance parameters are illustrated in Fig. 3, along with the corresponding Otsu's thresholding results, for two volunteers. We can clearly observe on the normal-weight volunteer the effect of the breathing on the localisation of the capsule interface, while we don't observe such effect on the obese volunteer. On the normal-weight volunteer, with higher temporal variability on SCD values, the average measures were the following: SCD_{IBS} = 12.08 ± 0.62 mm, SCD_{σ^2}=13.76 ± 0.52 mm, while the B-mode measure was $SCD_{B\text{-}mode}$ = 13 mm. Maximum and minimum SCD values over the temporal series

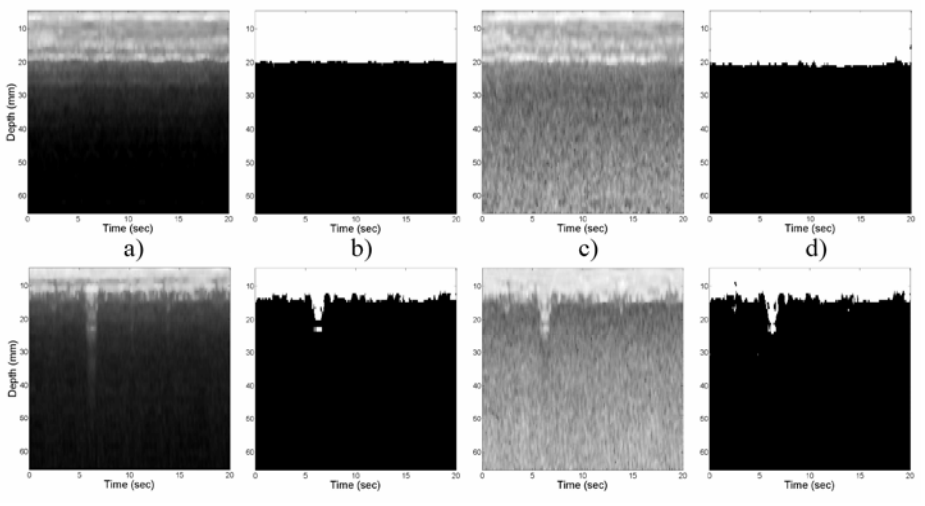


Fig. 3. Spectral analysis on an obese volunteer (top row) and a normal weight volunteer (bottom row). Normalized parametric maps of the IBS (a) the spectral variance (c) are plotted along with the corresponding thresholding results (b-d).

were: SCD_{IBS}=[11.07 mm 13.09 mm] and SCD_{σ^2} =[12.41 mm, 14.43 mm]. This corresponds to lateral liver displacement, relative to the skin surface, of 2mm which correlates well with MRI-based findings reported in [11].

3.2 Quantitative Evaluation of the SCD Measures

SCD measures based on both spectral parameters were first evaluated on an *in vitro* experiment to measure the layer thickness on subcutaneous pig tissues, as illustrated in Fig. 4. A pig meat sample made of fat and muscle was placed on top of a gelatin cylinder phantom mimicking the liver acoustic properties.

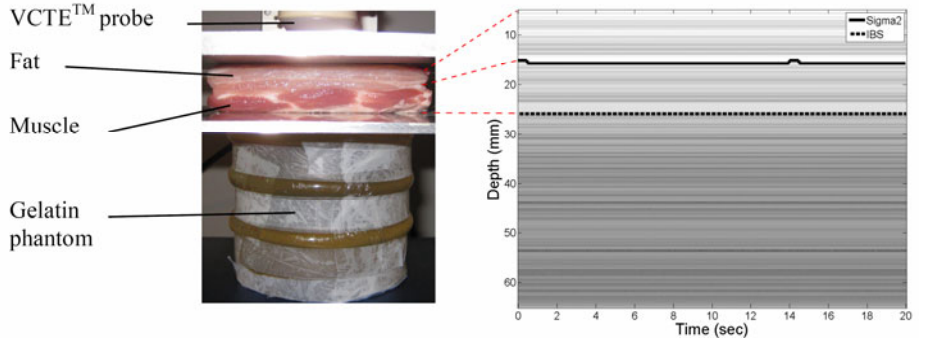


Fig. 4. In vitro experiment on pig tissues. (a) Experimental scanning setup. (b) Temporal series of RF lines. The black line on the image represents the SCD values calculated with the s^2 parameter (solid) and the IBS parameter (dashed).

The subcutaneous fat layer thickness directly measured on the sample was 13.4 mm. The algorithm provided a temporal average measure of 15 mm for the σ^2 parameter and 25mm for the IBS parameter, both estimations having a spatial precision of 1.25 mm (for the algorithm). This experiment, performed on high-quality RF lines, showed that the IBS parameter completely failed to provide an accurate fat layer thickness measure, while the σ^2 parameter provided an accurate estimate, within the precision of the method. This experiment highlights the superiority of the spectral variance parameter to distinguish fat from muscle and to be less sensitive to depth-attenuation which is the predominant advantage on *in vivo* RF data.

3.3 Qualitative Comparison of SCD Measures on Volunteers

SCD values measured with the three probes (1 B-mode and 2 VCTE™), and averaged over the whole population or separately over the normal-weight ($SCD_{B\text{-}mode}$<20 mm) and the over-weight population ($SCD_{B\text{-}mode}$>20 mm), are reported in Table 1. For the spectral estimations, the algorithm precision of 3λ corresponded to 1.8mm for the 2.5 MHz probe and 1.3mm for the 3.5 MHz probe. As expected, the SCD measures were smaller with VCTE™ probes which apply more pressure on the skin. The range of differences confirmed that $SCD_{B\text{-}mode}$ cannot be directly mapped to setup the scanning protocol with the VCTE™ probes. The two VCTE™ probes should provide similar SCD values, which was only the case with the σ^2 parameter. The IBS parameter estimated systematically lower SCD values with up to 3.7 mm (30%) of measure discrepancies between probes, for the over-weight population.

Focusing on the normal-weight population (SCD<20 mm), the differences between the B-mode and the VCTE™-based measures should be small, given the little deformability of the subcutaneous tissues. Regarding this point, we observed that the similarity of measures was better for the σ^2 parameter than for the IBS. We also observed that the IBS parameter systematically underestimated the subcutaneous layer thickness, especially for the 3.5 MHz probe. Finally, we observed that higher correlations (c.f. p-values) were obtained for the σ^2 parameter, especially for over-weight volunteers. These results overall suggested that this parameter, even though more costly to compute, should be preferred to the IBS parameter.

To evaluate the consistency between SCD measures from the B-mode and the VCTE™ probes, we performed Kendall's correlation tests on the whole database. Results are reported in Table 2, for the following parameters of the test: τ is the Kendall's rank correlation coefficient, and p-val is the p-value for the correlation test. The p-values confirmed that the SCD measures provided by the spectral methods were well correlated with B-mode ultrasound measures, while direct comparison was biased, especially on over-weight patients.

Table 1. Average SCD values measured with three probes

SCD (mm)	B-mode	VCTE™ 2.5 MHz		VCTE™ 3.5 MHz	
		σ^2	IBS	σ^2	IBS
All	17.2±5.9	16.0±4.3	12.2±3.9	15.6±4.3	10.7±2.6
<20	14.4±3.7	14.7±3.2	11.1±2.8	14.5±3.7	10.2±2.3
>20	25.1±6.1	20.4±4.6	15.9±4.8	19.4±4.1	12.2±3.0

Table 2. Kendall's correlation test between B-mode and VCTE™ SCD measures

Probe		All SCD		$SCD_{B\text{-}mode} \leq 20mm$		$SCD_{B\text{-}mode} > 20mm$	
2.5 MHz	σ^2	$\tau = 0.6$	p-val = 2.10^{-24}	$\tau = 0.5$	p-val = 6.10^{-15}	$\tau = 0.4$	p-val = 2.10^{-3}
	IBS	$\tau = 0.4$	p-val = 6.10^{-16}	$\tau = 0.3$	p-val = 1.10^{-7}	$\tau = 0.3$	p-val = 9.10^{-3}
3.5 MHz	σ^2	$\tau = 0.5$	p-val = 3.10^{-18}	$\tau = 0.4$	p-val = 6.10^{-12}	$\tau = 0.2$	p-val = 5.10^{-3}
	IBS	$\tau = 0.3$	p-val = 2.10^{-9}	$\tau = 0.3$	p-val = 1.10^{-6}	$\tau = 0.1$	p-val = 0.19

4 Conclusion

This paper presented the evaluation of a spectral analysis approach to segment ultrasound RF lines and measure subcutaneous layer thickness, for liver scanning, on a large database of subjects. A correlation study was performed to compare thickness measures with an imaging B-mode ultrasound probe. The spectral analysis was tailored to the specific scanning and acquisition setup of a VCTE™ probe, involving the tuning of the spatial window length, the weighting window and the spectral bandwidth. *In vitro* experiment on pig tissues reported a precision of measure of 1 to 2 mm.

References

1. Sandrin, L., Fourquet, B., Hasquenoph, J.-M., Yon, S., Fournier, C., Mal, F., Christidis, C., Ziol, M., Poulet, B., Kazemi, F.: Transient elastography: A new noninvasive method for assessment of hepatic fibrosis. Ultrasound in Medicine and Biology 29(12), 1705–1713 (2003)
2. Castera, L., Vergniol, J., Foucher, J., Bail, B.L., Chanteloup, E., Haaser, M., Darriet, M., Couzigou, P., Ledinghen, V.d.: Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. Gastroenterology 128(2), 343–350 (2005)
3. Hammoude, A.: Edge detection in ultrasound images based on differential tissue attenuation rates. Ultrasonic Imaging 21(1), 31–42 (1999)
4. Boukerroui, D., Basset, O., Baskurt, A., Gimenez, G.: A multiparametric and multiresolution segmentation algorithm of 3-d ultrasonic data. IEEE Transactions on Ultrasononics, Ferroelectrics and Frequency Control 1(48), 64–77 (2001)
5. Dydenko, I., Friboulet, D., Gorce, J.-M., D'hooge, J., Bijnens, B., Magnin, I.E.: Towards ultrasound cardiac image segmentation based on the radiofrequency signal. Medical Image Analysis 7(3), 353–367 (2003)
6. Moradi, M., Mousavi, P., Abolmaesumi, P.: Tissue characterization using fractal dimension of high frequency ultrasound RF time series. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 900–908. Springer, Heidelberg (2007)
7. Davignon, F., Deprez, J.F., Basset, O.: A parametric imaging approach for the segmentation of ultrasound data. Ultrasonics 43(10), 789–801 (2005)
8. Ng, J., Rohling, R., Lawrence, P.D.: Automatic measurement of human subcutaneous fat with ultrasound. IEEE Transactions on Ultrasononics, Ferroelectrics and Frequency Control 56(8), 1642–1653 (2009)
9. Huisman, H.J., Thijssen, J.M.: Precision and accuracy of acoustospectrographic parameters. Ultrasound in Medicine & Biology 22(7), 855–871 (1996)
10. Fink, M., Hotteir, F., Cardoso, J.F.: Ultrasonic signal processing for in vivo attenuation measurement: Short time Fourier analysis. Ultrasonic Imaging 5(2), 117–135 (1983)
11. Rohlfing, T., Maurer Jr., C.R., O'Dell, W.G., Zhong, J.: Modeling liver motion and deformation during the respiratory cycle using intensity-based free-form registration of gated MR images. In: SPIE Medical Imaging, San Diego, CA, USA (2004)

Incremental Shape Statistics Learning for Prostate Tracking in TRUS

Pingkun Yan and Jochen Kruecker

Philips Research North America
345 Scarborough Road, Briarcliff Manor, NY 10510

Abstract. Automatic delineation of the prostate boundary in transrectal ultrasound (TRUS) can play a key role in image-guided prostate intervention. However, it is a very challenging task for several reasons, especially due to the large variation of the prostate shape from the base to the apex. To deal with the problem, a new method for incrementally learning the patient-specific local shape statistics is proposed in this paper to help achieve robust and accurate boundary delineation over the entire prostate gland. The proposed method is fast and memory efficient in that new shapes can be merged into the shape statistics without recomputing using all the training shapes, which makes it suitable for use in real-time interventional applications. In our work, the learned shape statistics is incorporated into a modified sequential inference model for tracking the prostate boundary. Experimental results show that the proposed method is more robust and accurate than the active shape model using global population-based shape statistics in delineating the prostate boundary in TRUS.

1 Introduction

Transrectal ultrasound (TRUS) is currently the most commonly used imaging modality for image-guided biopsy and therapy of prostate cancer due to its real-time nature, cost effectiveness, and simplicity. Accurate delineation of the prostate boundary in TRUS can play a key role in image-guided prostate interventions. However, extracting the prostate boundary in TRUS is a challenging task due to the low signal-to-noise ratio of ultrasound imaging, the large intensity variation both inside and outside of the prostate, and especially the large shape variations through the whole gland from the base to the apex.

Methods for segmenting the prostate in static TRUS images have been reported in the past. Abolmaesumi and Sirouspour [1] traced the prostate boundary by moving from one estimated boundary point to the next following a trajectory described by a constant velocity dynamic model. Ladak et al. [2] employed a discrete dynamic contour guided by shape statistics to segment the prostate in TRUS. Shen et al. [3] proposed a segmentation method by first classifying a TRUS image using support vector machine with Gabor filter bank and then fitting a statistical prostate shape model to the output image. These static image segmentation methods have been successful in the mid-gland, because the

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 42–49, 2010.

boundary shapes in that area are quite regular with small variation. However, due to the large prostate shape variation in the base and the apex areas, poor segmentation performance in those areas was indicated [3,4]. Recently, Yan et al. [5] addressed the problem using a dynamic shape modeling approach. The prostate shape statistics is continuously recomputed from several of the most recently segmented prostate shapes. The method obtained considerable success by incorporating the adaptive local shape statistics into the segmentation process. However, due to the continuous model recalculation using the obtained shapes, this method is not computationally efficient and also requires storing all the training shapes.

In this paper, we propose a new method for incrementally learning the patient-specific local shape statistics to help achieve accurate boundary delineation over the whole prostate gland in TRUS. New shapes that become available during the delineation process can be merged into the shape statistics using an incremental subspace learning method [6,7] without explicitly recomputing using all the training shapes. Therefore, the proposed learning method is computationally efficient and removes the need of storing all the training shapes, which makes it suitable for use in interventional applications. In addition, the incremental learning method updates the shape statistics in a smoother way compared to the work in [5]. Instead of segmenting each frame separately, the shape statistics is incorporated into a modified sequential inference model for contour based tracking to exploit the shape change dynamics from frame to frame. The prostate boundary tracker starts with the global population-based shape statistics. As more and more prostate boundary shapes from a particular patient are available along with the tracking, the learned shape statistics will become patient-specific and can quickly adapt to the prostate boundaries in a local neighborhood by weighing more on the most recent shapes. Finally, since the proposed method does not employ any numerical optimization, it is much faster compared to the previous methods on prostate boundary delineation [2,3,5], which is a highly desired feature for use in real-time interventional applications. The performance of the method is demonstrated by tracking the prostate boundaries in a series of TRUS sequences.

2 Incremental Shape Statistics Learning

Let a prostate contour S_t in the frame acquired at time t represented by a series of k contour points, i.e. $S_t = \{\mathbf{v}_t^i | i = 1, \ldots, k\}$. Each contour point $\mathbf{v}_t^i$ is a 2D vector $\{x_t^i, y_t^i\}$. The contour points are equally spaced (Euclidean distance based spacing) and sampled from the shapes [5,9]. The active shape model (ASM) introduced by Cootes et al. [8] is used in our work for shape modeling, which provides a compact statistical shape representation by modeling contour point distribution. Suppose that the shape statistics has been computed from previous n prostate boundary observations at time t_0. The shapes can be denoted by $A = \{S_i - \bar{S}_A | i = 1, \ldots, n\}$, where $\bar{S}_A = \frac{1}{n}\sum_{i=1}^{n} S_i$ is the mean shape. The observed shapes can be decomposed by using the singular value decomposition

(SVD) as $A = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, where each column of $\mathbf{U}$ is an eigenvector of A and $\mathbf{\Sigma}$ is a diagonal matrix containing the corresponding eigenvalues. With the computed shape statistics, a new shape S can be represented by using the mean shape and the eigenvectors linearly combined by a coefficient vector $\mathbf{b} = \mathbf{U}^T(S - \bar{S}_A)$. The approximation of the shape S constrained by the shape statistics can be obtained by $\hat{S} = \bar{S}_A + \mathbf{U}\mathbf{b}$.

As the TRUS probe moves to image different areas of the prostate gland, it is desirable to use the newly available shapes to update the previous shape statistics to have the patient-specific local prostate shape knowledge. Let $B = \{S_i - \bar{S}_B | i = 1, \ldots, m\}$ represent the differences between the recent m shapes acquired after t_0 and their mean shape $\bar{S}_B = \frac{1}{m}\sum_{i=1}^{m} S_i$. The objective is to merge the new shape set B into the shape statistics computed from the set A at time t_0. To achieve this goal, the incremental subspace learning algorithm [6,7], which was originally used for learning the appearance of a rectangular area in video frames, is modified to incrementally learn the shape statistics of deformable contours. Thus, the shape statistics can be updated by merging the new observations without recomputing using all the training shapes. Let C denote the concatenation of shape sets A and B. The mean shape in the new shape statistics is computed as

$$\bar{S}_C = \frac{nf}{nf+m}\bar{S}_A + \frac{m}{nf+m}\bar{S}_B, \tag{1}$$

where f is a forgetting factor $f \in \mathcal{R}$ and $0 \leq f \leq 1$. The smaller the value of f, the faster the information from the old shape statistics is discarded, and vice versa. To assist the effective SVD computation, let $\tilde{B}$ be the component of B orthogonal to $\mathbf{U}$ and $\hat{B} = \left[B_1 \ \cdots \ B_m \ \sqrt{\frac{nm}{n+m}}(\bar{S}_B - \bar{S}_A)\right]$. The concatenated matrix C can be expressed as $C = [A \ B] = [\mathbf{U} \ \tilde{B}]\mathbf{R}\begin{bmatrix}\mathbf{V}^T & 0 \\ 0 & \mathbf{I}\end{bmatrix}$, where $\mathbf{R} = \begin{bmatrix} f\mathbf{\Sigma} & \mathbf{U}^T\hat{B} \\ 0 & \tilde{B}(\hat{B} - \mathbf{U}\mathbf{U}^T\hat{B})\end{bmatrix}$. Finally, the incrementally updated shape statistics of C can be obtained as $\mathbf{U}' = [\mathbf{U} \ \tilde{B}]\tilde{\mathbf{U}}$ and $\mathbf{\Sigma}' = \tilde{\mathbf{\Sigma}}$, where $\tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T$ is the SVD of $\mathbf{R}$.

The proposed incremental learning algorithm involves computation of QR decomposition of a $k \times (m+1)$ matrix and SVD of a $(e+m+1) \times (e+m+1)$ matrix for each update, where e is the number of eigenshapes at t_0. Compared to the simplistic approach of applying SVD to a $k \times (n+m)$ matrix for recomputing shape statistics at each step, our proposed method learns the shape statistics in a more efficient way, especially when the number of training shapes n gets large. There is also no need to store the $k \times n$ contour points of the old training shapes. The resulted reduction in computational time and memory usage makes the algorithm well-suited for real-time image-guidance tasks.

3 Sequential Inference Model for Tracking

Instead of separately segmenting each TRUS frame, the incremental shape statistics learning algorithm is incorporated into a visual tracking framework for

prostate boundary delineation to exploit the gradual shape transition along with the movement of the TRUS probe. Given a set of observed TRUS frames $\mathcal{I}_t = \{I_i | i = 1, \ldots, t\}$, the aim is to delineate the prostate shape S_t in the current frame t, which can be formulated as to find S_t that maximizes the probability $p(S_t|\mathcal{I}_t)$. By using Bayes' theorem, we have

$$p(S_t|\mathcal{I}_t) \propto p(I_t|S_t) \int p(S_t|S_{t-1}) p(S_{t-1}|\mathcal{I}_{t-1}) dS_{t-1}. \tag{2}$$

The probability $p(S_{t-1}|\mathcal{I}_{t-1})$ denotes the tracking result obtained in the previous frame. The probability $p(S_t|S_{t-1})$ describes the dynamic process of propagating the prostate shapes from frame $t-1$ to frame t. The probability $p(I_t|S_t)$ measures how well the propagated contour S_t can be fit into the frame t. The details of the probability terms on the right side of (2) are presented as follows.

For computational efficiency and robustness, prostate shapes are projected into the subspace of the incrementally learned shape statistics by using an affine transformation $\mathbf{T}_t$ and the coefficient vector $\mathbf{b}_t$ as in Section 2

$$S_t = \mathbf{T}_t \left(\bar{S}_t + \mathbf{U}_t \mathbf{b}_t \right). \tag{3}$$

Let $\mathbf{x}_t = [x_t, y_t, \theta_t, s_t, \alpha_t, \phi_t]^T$ be the translations along x and y axes, rotation angle, scale, aspect ratio, and skew direction in the affine transformation. The state vector $\mathbf{X}_t$ can be composed by concatenating the transformation parameters and shape representation coefficients as $\mathbf{X}_t = [\mathbf{x}_t; \mathbf{b}_t]$. The prostate shape transition between the TRUS frames can be described as

$$\mathbf{X}_t = \mathbf{X}_{t-1} + \Delta\mathbf{X} = [(\mathbf{x}_t + \Delta\mathbf{x}); (\mathbf{b}_t + \Delta\mathbf{b})], \tag{4}$$

where the difference $\Delta\mathbf{X}$ follows normal distribution $\mathcal{N}(\Delta\mathbf{X}; 0, \mathbf{\Psi})$ and $\mathbf{\Psi}$ is a diagonal covariance matrix whose elements are the corresponding variances of the elements in $\mathbf{X}_t$. Therefore, we have

$$p(S_t|S_{t-1}) = p(\mathbf{X}_t|\mathbf{X}_{t-1}) = \mathcal{N}(\mathbf{X}_t; \mathbf{X}_{t-1}, \mathbf{\Psi}). \tag{5}$$

The probability $p(I_t|S_t)$ of observing TRUS frame I_t given a shape S_t can be measured by the distance between S_t and the detected prostate boundary. The dark-to-bright transition has been shown to be an effective prostate boundary indication feature [9], which can be computed by the pixel intensity contrast along the normal vector of the contours. Let $\{F^i_{max} | i = 1, \ldots, k\}$ denote the points with the maximal feature values corresponding to each contour point, the probability $p(I_t|S_t)$ can be measured

$$p(I_t|S_t) \propto \exp\left(-\frac{1}{2r^2 k} \sum_{i=1}^{k} \|F^i_{max} - x^i_t\|^2 \right), \tag{6}$$

where r is the spacing between the points on the normal vector profile.

The particle filter based CONDENSATION algorithm introduced by Isard and Blake [10] is modified for solving this deformable contour tracking problem formulated in (2). The workflow of the tracking method is shown in Fig. 1. Since the

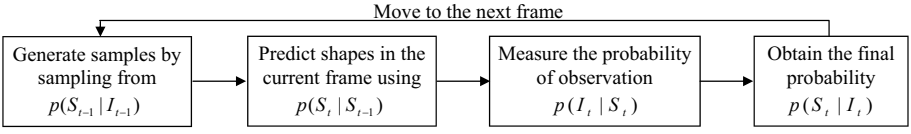


Fig. 1. Workflow of the particle filter based tracking method using the sequential inference model

proposed method does not employ any numerical optimization, it runs in linear time $O(N)$, where N is the number of samples. The final shape can be estimated by computing the expectation of all the samples $\hat{S}_t = \sum_{i=1}^{N} p(S_t^i|\mathcal{I}_t)S_t^i$.

4 Experimental Results

In our experiments, the proposed method was retrospectively tested on sequences collected during TRUS guided prostate cancer biopsy procedures using an iU22 ultrasound system (Philips Healthcare, Andover, MA). The TRUS frames were digitized by a video card at 30 frames per second (fps). Each grabbed TRUS frame has 640×480 pixels and the pixel sizes were 0.1493mm and 0.1798mm in 4cm and 5cm depth settings, respectively. In total, 3215 video frames were grabbed from 20 different patients. An experienced radiologist manually delineated the prostate boundary in one out of every ten frames of the TRUS sequences, in total 316 frames, to provide the ground truth for validation.

In our experiments, the tracked prostate boundary was represented by a contour with $k = 64$ points. The initial prostate shape statistics was computed using 62 prostate shapes previously obtained from a different data set. To quickly incorporate the newly tracked shapes into the shape statistics, the incremental learning was performed in every frame. The forgetting factor f in (1) was empirically set to 0.95 to weigh the learning more on the recently tracked contours, while maintaining the smoothness of the incrementally learned shape statistics. The number of samples N used in the particle filter is a trade-off between the efficiency and accuracy. A larger number of samples may give a bit more accurate tracking results, however, will also make the algorithm slower. The results demonstrated in this paper were obtained by using N=1000 samples.

The proposed method starts tracking from a given initial contour, which can be obtained by either manual delineation or using an automatic 2D TRUS image segmentation algorithm [3,5]. The initial values of the affine transformation parameters and the shape representation coefficients can be computed by using the initial contour. The variance of the normal distribution in the dynamical model (5) was empirically set to be $[2.0, 2.0, 0.001, 0.05, 0.05, 0.001]$ for the transformation part and $\sqrt{\Sigma}/10$ for the shape representation part. Only the first 3 most significant eigenvectors in the shape statistics were used in our work. The proposed method was developed in Matlab. Once initialized, it tracks the prostate boundary in TRUS fully automatically. On a Core2 1.86 GHz PC, we achieved

Table 1. MAD errors (mm) of the ASM and the proposed tracking method on delineating the prostate boundary in the TRUS sequences

Method	seq-1	seq-2	seq-3	seq-4	seq-5	seq-6	seq-7	seq-8	seq-9	seq-10	
ASM	3.28	2.54	2.24	1.66	3.09	2.54	1.73	2.67	2.34	2.80	
Tracking	0.95	1.92	1.45	1.62	2.16	0.83	0.63	0.76	1.54	1.36	
	seq-11	seq-12	seq-13	seq-14	seq-15	seq-16	seq-17	seq-18	seq-19	seq-20	Average
ASM	1.66	1.20	3.01	1.63	3.02	4.73	1.92	2.97	2.53	2.61	2.51±0.78
Tracking	1.52	0.96	1.89	1.46	2.26	1.22	1.49	1.98	1.42	1.39	1.44±0.46

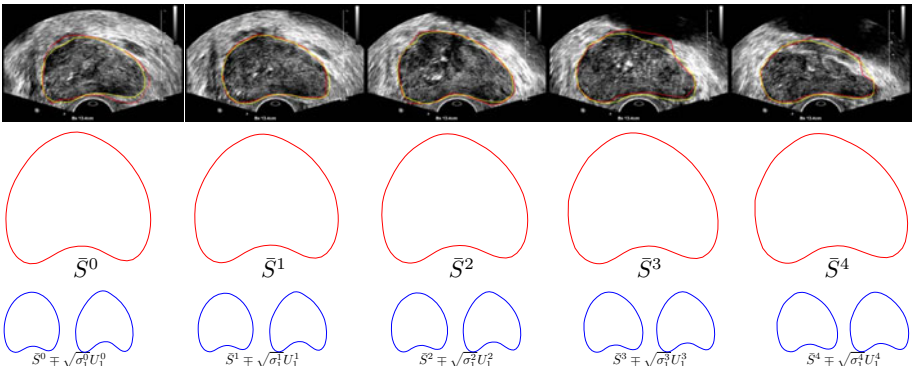


Fig. 2. The 1st row shows the delineation of prostate boundary in TRUS with the manually delineated contour in red and the automatic tracking result in yellow; the 2nd row shows the mean shapes in the incrementally learned shape statistics at each stage corresponding to the video frames; the 3rd row shows the shapes generated by varying the mean shapes with the largest mode of variation

12 fps with the above parameter setting. On the same computer, the ASM segmented the sequences with 3 fps.

Fig. 2 shows the tracking results of a TRUS sequence and the learned shape statistics corresponding to the displayed frames. It can be observed that the mean shapes and the major shape components gradually change towards the tracked shapes through incremental learning. Fig. 3 demonstrates the tracking results of the proposed method in different regions of the prostate gland. The initialization contours are included to show that the proposed method can successfully track the prostate boundary when the prostate boundary shape becomes significantly different from the initial shape. In general, the tracking results are close to the manually delineated contours.

Besides the visual inspection, we also quantitatively evaluated the performance of the proposed method. The ASM [8], which has been widely used for segmenting the prostate in TRUS [2,3,5,9], was applied to the same data set by detecting the prostate boundary using the same boundary feature and initial contours for comparison. For TRUS sequence segmentation, the result of ASM

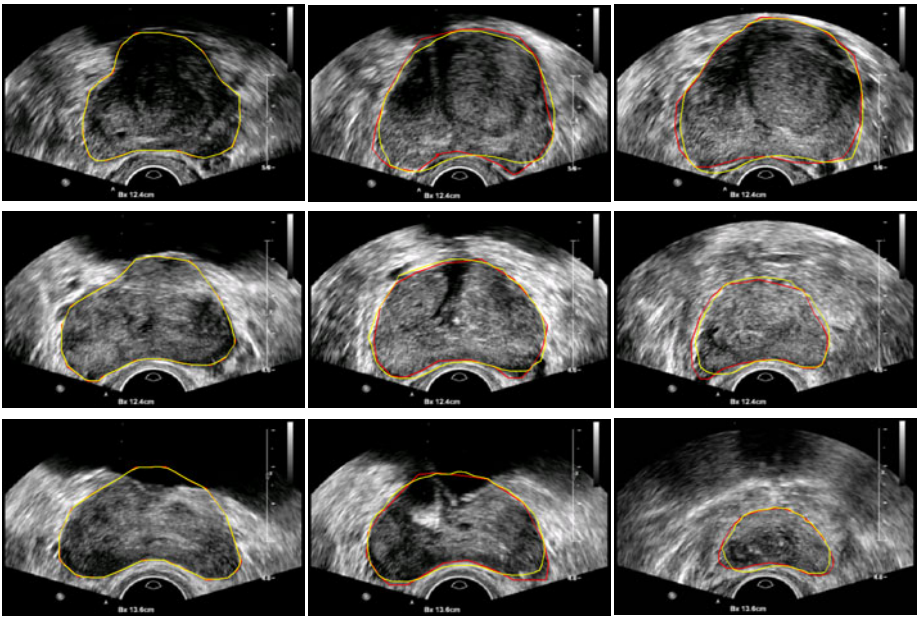


Fig. 3. The first column shows the initialized contours to start the tracking of each sequence. The rest two columns show the tracking results in different part of the prostate gland, where the manually delineated contours are shown in red and the tracked contours are shown in yellow.

Table 2. MAD errors (mm) of the ASM and the proposed tracking method in the base, mid-gland, and apex areas over the 20 TRUS sequences, respectively

Method	base	mid-gland	apex
ASM	2.53±1.17	1.48±0.59	3.52±1.63
Tracking	1.39±0.67	1.20±0.42	1.73±0.76

from one frame was propagated to the next as the initialization. The mean absolute distance (MAD) error was employed for quantitative measurement. As shown in Table 1, the accuracy of the proposed method was significantly better than the ASM, which uses only the global population-based shape statistics. In our work, the statistical significance was evaluated using paired t test ($p < 0.05$).

It is also interesting to see the performance of the methods in each of the base, mid-gland, and apex regions. For quantitative evaluation, the ground truth segmentations of each patient were sequentially divided into three groups with equal number of frames in each group. Since the sequence was obtained by scanning through the prostate from the base to apex, the three groups were labeled as base, mid-gland, and apex in the specified order. The evaluation results are shown in Table 2. Compared to the ASM, prostate boundary delineation

accuracy was significantly improved in all the regions by using the proposed tracking approach with incrementally learning the shape statistics.

5 Conclusions

In this paper, we presented a new algorithm for incremental shape statistics learning, which is well-suited for use in image-guidance applications due to its computation and memory efficiency. The learning algorithm was successfully incorporated into a modified sequential inference model to help achieve more robust and accurate delineation of the prostate boundary in TRUS over the entire prostate gland. Experimental results show that the contour delineation error of the proposed method was 42.6% less than that of the ASM. Significant improvement was made in the base and apex areas with 45.1% and 50.1% less delineation errors, respectively. To the best of our knowledge, the proposed tracking algorithm is the first method that aims to extract the prostate boundary from TRUS in real-time during image-guided prostate interventions. In our future work, we will investigate more efficient software implementation of the algorithm and also explore hardware acceleration to get the program run in real-time.

References

1. Abolmaesumi, P., Sirouspour, M.R.: An interacting multiple model probabilistic data association filter for cavity boundary extraction from ultrasound images. IEEE Trans. Med. Imaging 23(6), 772–784 (2004)
2. Ladak, et al.: Prostate boundary segmentation from 2D ultrasound images. Med. Phys. 27(8), 1777–1788 (2000)
3. Shen, et al.: Segmentation of prostate boundaries from ultrasound images using statistical shape model. IEEE Trans. Med. Imaging 22, 539–551 (2003)
4. Yan, et al.: Optimal search guided by partial active shape model for prostate segmentation in TRUS images. In: Medical Imaging 2009, vol. 7261. SPIE (2009)
5. Yan, et al.: Segmenting TRUS video sequences using local shape statistics. In: Medical Imaging 2010, vol. 7625. SPIE (2010)
6. Levy, A., Lindenbaum, M.: Sequential karhunen-loeve basis extraction and its application to images. IEEE Transactions on Image Processing 9(8), 1371–1374 (2000)
7. Ross, D.A., Lim, J., Lin, R.S., Yang, M.H.: Incremental learning for robust visual tracking. International Journal of Computer Vision 77, 125–141 (2008)
8. Cootes, et al.: Active shape models – their training and application. Comput. Vision Image Understand. 61(1), 38–59 (1995)
9. Hodge, et al.: Prostate boundary segmentation from ultrasound images using 2D active shape models. Comp. Methods Prog. Biomed. 84, 99–113 (2006)
10. Isard, M., Blake, A.: CONDENSATION – conditional density propagation for visual tracking. International Journal of Computer Vision 29(1), 5–28 (1998)

Fast and Accurate Ultrasonography for Visceral Fat Measurement

You Zhou[1], Norihiro Koizumi[1,2], Naoto Kubota[2,3], Takaharu Asano[3], Kazuhito Yuhashi[3], Takashi Mochizuki[4], Takashi Kadowaki[2,3], Ichiro Sakuma[1,2], and Hongen Liao[1,2]

[1] Graduate School of Engineering, The University of Tokyo
[2] Translational Systems Biology and Medicine Initiative, The University of Tokyo
[3] Graduate School of Medicine, The University of Tokyo
[4] Aloka Co. Ltd., Japan
{zhouyou,sakuma,liao}@bmpe.t.u-tokyo.ac.jp

Abstract. Visceral fat area (VFA) has close relationship with hypertension, diabetes and cardiovascular disease, and therefore serve as a reliable indicator of these diseases. Abdominal computed tomography (CT) enables precise quantification of the VFA and has been considered as the gold standard for VFA assessment. In this paper, we develope a novel method to quickly and accurately measure the VFA with ultrasonography (US). We evaluated the novel method on five volunteers and the diagnosis procedures lasted less than 30 seconds averagely. The simulation results by our method were compared with VFA estimated by abdominal CT. The correlation coefficient between them was 0.913 for men and 0.858 for women. And the mean deviation of between VFA by CT and by our method was $19.8\ cm^2$ for men and $13.3\ cm^2$ for women.

1 Introduction

The term metabolic syndrome (MS), a combination of medical disorders that increase the risk of chronic diseases such as diabetes, hypertension, and cardiovascular disease, has been adopted by International Diabetes Federation (IDF) in 2006. MS is a major public health problem, the prevalence of which has increased worldwide. Visceral obesity, the excess accumulation of visceral fat deposits mainly around the waist, is thought to be a fundamental pathology for MS in particular. Therefore, accurate measurement of visceral fat represents an important tool in assessing MS.

Visceral fat, also known as organ fat, packes in between internal organs and the torso, as opposed to subcutaneous fat which is found underneath the skin. Since visceral fat is located deep inside the body and mixed with other organs, its accurate assessment has prooven to be challenging work. Abdominal computed tomography (CT) has been considered the most accurate and reproducible technique of body fat measurement, particularly abdominal adipose tissue. Abdominal CT enables accurate quantification of the visceral fat area (VFA) and therefore serves as the gold standard for visceral fat assessment [1]. However, abdominal CT has many drawbacks, including exposure to radiation, lack of simplicity, high cost, and time-consuming.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 50–58, 2010.

Due to these limitations, a variety of alternative methods are being used to assess visceral fat amount and distribution. Previous studies have shown that waist circumference (WC), and WC-based indices can perform as some indicators of the level of visceral obesity [2]. These measurements are recommended as a simpler and easier screening method. However, due to the lack of the individual visualization of visceral fat, these methods have fatal drawbacks such as inability to distinguish subcutaneous fat from visceral fat, inability to take into account an individual's specific information, low level of reproducibility in the case of marked obesity, and most important, poor accuracy.

In recent years, simple methods for assessing visceral fat accumulation using ultrasonography (US) have been studied and were further confirmed by strong correlations with CT-detected VFA [3][4]. Ultrasound has many advantages such as non invasive to human bodies, low-cost, easy to be operated, real time and enables the visualization of visceral fat. Therefore, US-based methods show a balance on simplicity and accuracy between abdominal CT and WC-based indices. However, due to the poor image quality and limited visual field of US, the results of the existing US-based methods are still unsatisfactory and need further improvements.

In the study, a fast and accurate method is developed for visceral obesity studies to provided an estimation of the abdominal visceral fat area and distribution using ultrasonography.

2 Method

2.1 Ultrasound Probe Compatible Device

To provide a quick, easy-operated and accurate way to guide the ultrasonographic procedures, three factors are important: 1) Same positions and angles for ultrasound probe on different patients during diagnosis; 2) Distinct markers easy to be observable in US image; 3) Quantitive measurement on patients' curvature of abdomen.

We designed a belt-shaped ultrasound probe compatible device to fulfill these requirements (**Fig.1**). Two kinds of US measurements of visceral fat are taken. Firstly, US-determined visceral fat distance is defined as the distance between the internal face of the rectus abdominis muscle and the centre of the aorta from each diagnostic position. Secondly, the images acquire from the central position were saved for further processing.

Unlike CT, the ultrasound can only provide a limited scope inside one patient's body. The diagnosing angle and position of probe may differ from doctor to doctor due to their own experiences. This belt-shaped device provides doctors a standard which is easy to follow. In the visual field of an abdominal US, due to the flow of blood, the aorta is nearly the most distinct marker for US. Thus, the belt-shaped device is designed to be fixed in the exact position from where the aorta can be observed most clearly. As a result, the diagnosis process will be finished in seconds, which is convenient for patients as well. With an elastic material, this belt-shaped device would bend smoothly and mold to patient's abdomen. It is easy to understand that: a

patient with mild visceral obesity in normal range of waist circumference would have a 'flatter' shape in abdomen, which means that the belt-shaped device will less bend when diagnosing, while a patient with more serious visceral obesity would get a 'plumper' waist, and resultingly more bent the belt-shaped device becomes when diagnosing. Hence, the curvature of the belt is proportional to the curvature of abdomen, which can serve as valuable characteristic of individual seriousness level of visceral obesity and give a reasonable classification of patients.With three probe-diagnosing positions rather than one, the curvature of abdomen can be described quantitively.

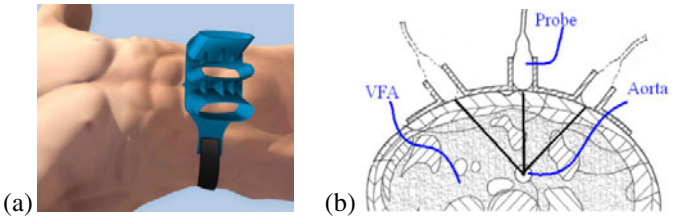


Fig. 1. Ultrasound probe compatible device. Distance and angle between two adjacent probes are 50mm and 40° in unbent situation. (a) Belt-shaped device design (b) Measurement method.

2.2 Segmentation of Visceral Fat Area

Analyzed results based on hundreds of abdominal CT images show that most visceral area of patient in his/her abdominal cross-section could be simulated as an ellipse, with aorta in the center. Although this approach may not be very exact for individual patient, the eccentricity of ellipse reflects the curvature of abdomen and therefore serves as a quantitive description of visceral obesity.

The ultrasound probe were detecting from three positions by the belt-shaped device. The information we can acquire contains: b, denotes the distance between the internal face of the rectus abdominis muscle and the centre of the aorta detected from front, and is defined as semi-minor axis of ellipse; b', denotes the distance between front-end of ultrasound probe in the central diagnosing position and the centre of the aorta detected from front, r, denotes the same distance detected from one side; h', denotes the distance between two probes (50mm); h, denotes the curve length between b and r derived from h' by $h = h'b/b'$as showed in **Fig. 2(a)**. Then, by ellipse circumference and ellipse parametric equation, we can calculate: a, denotes the semi-major axis; θ, denotes the angle between two probes.

For abdominal cross-section of a patient who has been diagnosed (denotes as patient$*$), we discretize the visceral area based on the ellipse approach as follow. The ellipse is divided uniformly by angle and axis length. Each section is identified by a number i. For patient$*$, we define the area of section i as a_i^*, the fat percentage of section i as p_i^*. Succinctly, we have $\boldsymbol{A}^* = (a_1^*, a_2^*, \dots, a_n^*)^T$ and $\boldsymbol{P}^* = (p_1^*, p_2^*, \dots, p_n^*)^T$, where n refers to the total number of sections (In this paper,

$n \triangleq 30 \times 40 = 1200$). Similar with a matrix, we can define N_i^*, the neighbourhood of p_i^* as: the fat percentages of sections who share the same edges or vertices with section i (**Fig.2(b)**). Immediately, there follows

$$VFA = \boldsymbol{A}^{*T}\boldsymbol{P}^* \quad . \tag{1}$$

The computation of $\boldsymbol{A}^*$is trivial and main difficulty is how to give a reasonable estimation of $\boldsymbol{P}^*$.

In order to distinguish the visceral fat from other organs, we used a Markov Random Field (MRF) based segmentation algorithm **(Fig. 3)**. The main process of this algorithm follows [5]. We must declare here that this kind of algorithms were not developed specially for solving the fat recognition problem of ultrasonographic image, and have never been introduced into this field.

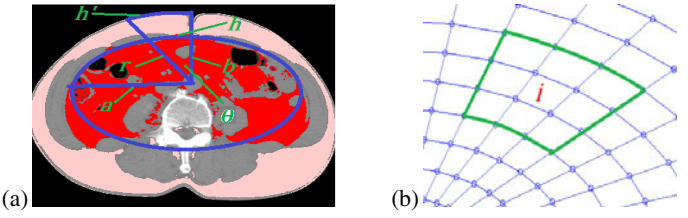


Fig. 2. Ellipse approach of abdominal cross-section. (a) Parameters of ellipse. Red part is the visceral fat. Grey part is other organs. (b) Neighbourhood in a discretized ellipse. Green part refers to N_i^*, the neighbourhood of p_i^*.

Fig. 3. Distinction of visceral fat. (a) Original US image. The colorful part in central indicates aorta. (b) Segmentation of visceral fat. The white part indicates the visceral fat area. (c) Matching US visual field into the whole visceral area.

2.3 Matching with Pre- stored Data

In this part, we compare the ultrasonographic image of patient$*$, with pre-stored CT scanned image of previous patients, and choose the patients whose situations are most similar with the current patient. We established a set V from 146 patients abdominal CT scanned images in a urban hospital, which cover full gamut of patients in its community. This pre-stored CT scanned images served as a pre-stored database with which the current patient$*$ can compare.

The visual field of ultrasonography is only local. So, we match the ultrasonography into the whole visceral area, and divide it into m sections following the previous part **(Fig. 3(c)).** We denote section i has area $a_i^*$and fat percentage p_i^*, for every $i = 1,2,\dots,m$. We have an estimation of m components of $\boldsymbol{P}^*$**,** and the problem now is how to estimate the fat percentage of the remaining $(n-m)$ hidden sections $\boldsymbol{P_h^*} = (p_{m+1}^*, p_{m+2}^*, \dots, p_n^*)^T$. For CT images, the discretization of ellipse can be accomplished without difficulty. Then, section i of patient k, has the fat percentage P_i^k. Naturally, there is $\boldsymbol{P}^k = (p_1^k, p_2^k, \dots, p_n^k)^T$, for $k = 1,2,\dots,146$.

Now, we want to find out the patients from V whose situations are most similar with patient$*$. As a result, the visceral fat distributions of those patients would serve as references for patient$*$. We evaluate the correlative level between patient$*$ and patient k by matching $p_1^*, p_2^*, \dots, p_m^*$ and $p_1^k, p_2^k, \dots, p_m^k$. We use the Grey-Level Correlation Formula to accomplish it

$$r^k = \frac{\sum_{i=1}^{m}(p_i^* - \overline{p^*})\left(p_i^k - \overline{p^k}\right)}{\sqrt{\sum_{i=1}^{m}(p_i^* - \overline{p^*})^2 \sum_{i=1}^{m}\left(p_i^k - \overline{p^k}\right)^2}}\ , \tag{2}$$

where $\overline{p^*}$ and $\overline{p^k}$ refer to the average of $p_1^*, p_2^*, \dots, p_m^*$ and $p_1^k, p_2^k, \dots, p_m^k$ respectively. Before the matching process, the patients whose eccentricity e and minor axis b differ too large from patient$*$ will be rejected in order to accelerate. The 10 patients with max value of r^k compose the set B. The CT image of those chosen patients would serve as the basis of fat estimation of current patient$*$.

We expect that $\boldsymbol{P_h^*}$ can be expressed as a 'combination' of patients from B and this combination can reflect the visceral fat distribution of patient$*$. To achieve it, we establish an optimization standard. By (**3**), we defines the cost at one section, and (**4**) is the integrally cost, which should be minimized. Consider $N_i^* \subset \boldsymbol{P}^*$, the neighbourhood of p_i^*. We give a choice not only realizing optimization at single p_i^* itself, but also throughout its neighbourhood N_i^*. We define the distance cost function between patient$*$ and patient k at section i as

$$\left(\left|p_i^* - p_j^*\right| - \left|p_i^k - p_j^k\right|\right)^2\ ,\ p_j^* \in N_i^*, p_j^k \in N_i^k\ , \tag{3}$$

where N_i^* and N_i^krefers to the neighbourhood of $p_j^*$and p_j^k respectively. Hence, integrally, the $\boldsymbol{P_h^*}$ should be the vector who satisfies that

$$arg\ \min_{p_i^* \in \boldsymbol{P_h^*}} \sum_{i=m+1}^{n} \sum_{k=1}^{10} \sum_{p_j^* \in N_i^*, p_j^k \in N_i^k} \left(\left|p_i^* - p_j^*\right| - \left|p_i^k - p_j^k\right|\right)^2, \tag{4}$$

$$p_i^* \in [0,1], \text{ for } i = 1,2,\dots,n.$$

Consequently, the determination of $\boldsymbol{P_h^*}$ has been transformed into an optimization problem. This numerical optimization is performed automatically by a constraint nonlinear conjugate gradient method (CGM).

3 Experiment Results

In this part, all correlations between two variables were evaluated using Pearson's product-moment correlation coefficient. Comparisons between each two groups were done with an unpaired t-test.

3.1 Evaluation Experiment of US

We tested the belt-shaped ultrasound probe compatible device on five volunteers from an urban community. All our ultrasonographic procedures were performed by the same examiner using a portable ultrasonography equipment (α10, ALOKA, Japan). The center of the belt-shaped device fixed exactly 2 cm left to umbilicus for each patient. Each patient assumed a supine position, and US data were measured at the end of expiration by a 6-MHz 2D abdominal probe, while the probe was making contact with patients' skin as slight as possible (**Fig.4**).

For each volunteer, the diagnosis procedures lasted less than 30 seconds. This is quite an acceptable time during a medical examination. In addition, More quick diagnosis speed is also expectable in the future if doctors become more proficient in this novel method.

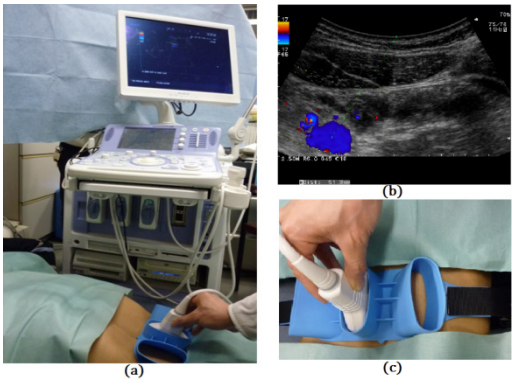


Fig. 4. Practical diagnosis procedures for a patient. (a) Diagnosis by US. (b) Distance of visceral fat measured in US image. (c) Belt-shaped ultrasound probe compatible device.

3.2 Campare with CT Results

Simulation on 146 patients' abdominal CT images were done and showed that there was strong correlation between the US observed VFA from front psotion and the total VFA of an individual patient. The coefficent was $r = 0.883$ and the level of significance was p-value < 2.2e-16 (**Fig. 5**). As a result, it is reasonable to use the US observed fat percentage as an indicator for total VFA.

Since abdominal CT holds the most accurate quantification of the VFA, we treated it as the testing standard for visceral fat assessment. Abdominal cross-section by CT scanning was obtained in a single tomographic slice at umbilical level as ultrasound probe did. Portions with a CT number of −200 to −10 Hounsfield Units (HU) were separated as adipose tissue and their areas were automatically calculated.

We simulated our method on 28 male and 35 female samples, and compared the results with abdominal CT. **Figure 6** shows the comparisons of CT and our US-based method. The mean deviation between VFA measured by CT and our US method for each patient is 19.8 cm^2 for men and 13.3cm^2 for women. The correlation between these two groups is also presented. For men, the coefficent was r = 0.913 and the level of significance was p-value = 1.179e-11. For women, there are r = 0.858 and p-value = 4.674e-11. These results proofed the significant positive correlations between VFA measured by CT and by our method.

And we also divided these 63 patients in both sex into three groups by the CT-measured VFA: low ($\leq 100cm^2$), medium (between $100cm^2$and $140cm^2$) and high ($> 140cm^2$). The average and max VFA deviations between US method and CT method of these three groups were shown in **Table 1.** From this table, it is easy to notice that our US-based method performed better on the medium group than other two groups averagely. The reason of this is that the data per-stored data mainly laid mainly in this VFA range. Thus, the result reminds us that the pre-stored database is very essential for our method and it must cover full range of patients in enough quantities. On the other hand, the max VFA deviations are all too large in three groups. This demonstrates that there are many personal detailed characteristics of certain patients have been omitted in our method, which should be checked carefully for the further improvement of this US-based method.

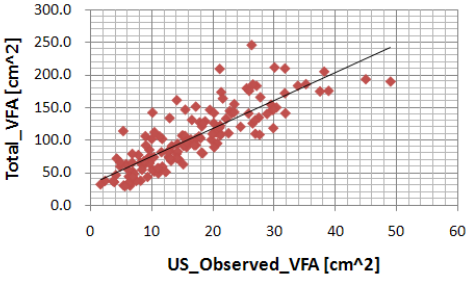


Fig. 5. Scattergram of relationship between US observed VFA and the total VFA

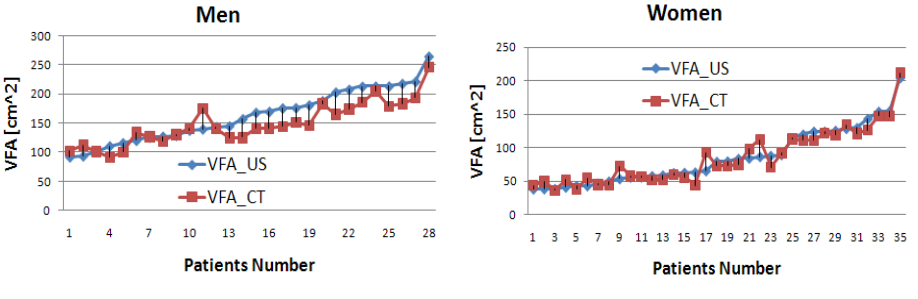


Fig. 6. Camprison of our US-based method and CT for men and women

Table 1. VFA deviations between US_method and CT_method in different groups

Groups		low	medium	high
Numbers of patients		17	25	21
VFA deviations (cm^2)	Average	15.7	9.1	27.3
	Max	42.2	36.9	37.3

4 Discussions and Conclusion

This paper proposes a novel US-based method for estimation of the VFA. A fast and convenient diagnosis method was introduced and a robust algorithm was proposed to estimate a patient's VFA from US data automatically. As far as we know, this is the first attempt to compute VFA by ultrasound, and our results were shown to be strongly correlated with CT measurements.

However, one main limitation of this present study is that all the data of patients were collected from the same ethnic group. Thus, the results cannot generalize directly to other individuals of certain ethnic groups. For instance, the female body type in particular differs largely between Westerners and Easterners. As a result, how to apply this novel method universally needs further discussions.

Till now, we have developed a US-based method whose accuracy reaches the level of abdominal CT. However, it is not an entirely impossible thing that the US-based method will become a more reliable indicator for MS than abdominal CT in the future. Due to the non-invasiveness and simplity of US, diagnosis at different levels of the abdomen will be accomplished without much difficulties. Consequencely, 3D individual visualization and estimation of visceral fat can be realized by US, while usually only one slice of abdominal cross-section image can be acquired by CT in medical examination due to the harmness of rediation. Secondly, due to its real-time characteristics, US allows us to detect visceral fat over a period of time. One patient's abdominal cross-sectional area when expiring will differ largely from inspiring. Abdominal CT cannot take this difference into consider and therefore would be less convincing than US in this respect. Further, by US, we may represent a useful method for monitoring weight loss, variations and transfer of visceral fat, which can be expected to indicate the associated risks of MS more accurately.

In conclusion, although its results are still rough and there are many details need discussing, the present US-based method has proofed to be a considerably fast and accurate way for VFA estimation.

Acknowledgment

This work was supported in part by Grant for Translational Systems Biology and Medicine Initiative (TSBMI) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

1. Rossner, S., Bo, W.J., Hiltbrandt, E., et al.: Adipose tissue determinations in cadavers—a comparison between cross-sectional planimetry and computed tomography. Int. J. Obes. Relat. Metab. Disord. 14, 893–902 (1990)
2. Hsieh, S.D., Yoshinaga, H., Muto, T.: Waist-to-height ratio, a simple and practical index for assessing central fat distribution and metabolic risk in Japanese men and women. Int. J. Obes. Relat. Metab. Disord. 27, 610–616 (2003)
3. Chiba, Y., Saitoh, S., et al.: Relationship between Visceral Fat and Cardiovascular Disease Risk Factors: The Tanno and Sobetsu Study. Hypertension Research 30, 229–236 (2007)
4. Ribeiro-Filho, F.F., et al.: Methods of Estimation of Visceral Fat: Advantages of Ultrasonography. Obesity Research 11(12), 1488–1494 (2003)
5. Lorenz, A., Haas, C., Ermert, H.: Segmentation of ultrasonic prostate images using a probabilistic model based on Markov random processes. Ultrason Imaging 19, 44–45 (1997)

Real-Time Gating of IVUS Sequences Based on Motion Blur Analysis: Method and Quantitative Validation

Carlo Gatta[1,2], Simone Balocco[2], Francesco Ciompi[2], Rayyan Hemetsberger[3], Oriol Rodriguez Leor[3], and Petia Radeva[2]

[1] Dept. Matemtica Aplicada i Anlisi, Universitat de Barcelona, Gran Via 585, 08007 Barcelona, Spain
[2] Centre de Visió per Computador, Bellaterra, Spain
[3] Unitat d'hemodinàmica cardìaca hospital universitari Germans Trias i Pujol Badalona, Spain

Abstract. Intravascular Ultrasound (IVUS) is an image-guiding technique for cardiovascular diagnostic, providing cross-sectional images of vessels. During the acquisition, the catheter is pulled back (pullback) at a constant speed in order to acquire spatially subsequent images of the artery. However, during this procedure, the heart twist produces a swinging fluctuation of the probe position along the vessel axis. In this paper we propose a real-time gating algorithm based on the analysis of *motion blur* variations during the IVUS sequence. Quantitative tests performed on an in-vitro ground truth data base shown that our method is superior to state of the art algorithms both in computational speed and accuracy.

1 Introduction

Intravascular Ultrasound is a catheter-based invasive diagnostic procedure used intraoperatively to assess cardiovascular diseases. The ultrasound probe is placed at the tip of the catheter and, while rotating on its axis, it emits and receives ultrasound pulses, successively reconstructed in radially distributed A-lines. The envelope detection of acquired radio frequency signals allows the reconstruction of cross-sectional images of the artery. During the acquisition, the catheter is pulled back at a constant speed in order to acquire spatially subsequent images of the artery. However, during this procedure, the heart twisting produces artificial fluctuations of the probe position along the vessel axis (swinging effect). Moreover, due to the heart cyclic contraction/expansion, an apparent rotation with respect to the catheter axis and in-plane translation can be observed.

The image visualization and automatic analysis, e.g. volumetric lumen segmentation and/or volumetric plaque characterization of IVUS pullbacks require a pre-processing motion compensation [1,2]. Given the intraoperative nature of IVUS imaging and the need of instantaneous diagnosis of the patient conditions, the motion compensation cannot be computed off-line and the computational cost of the gating algorithm must be very low. The combined effect of the three

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 59–67, 2010.

movements can be reduced by image-based gating algorithms [3,4,5] i.e. by extracting the most stable frames from the pullback sequence. Such techniques are composed by two parts: 1 - the extraction of a signal (1D or 2D) correlated with the cardiac phase; 2 - the selection of stable frames based on the previously extracted signal. The method in [3] extracts two signals based on two different principles; they conclude that the more reliable is based on the computation of the Absolute Intensity Difference (AID) applied to subsequent frames of a circular ROI centered in the middle of the short-axis image. The method in [4] computes a Dissimilarity Matrix based on the Normalized Cross Correlation (NCC) between couples of frames; subsequently a dynamic programming algorithm finds optimal gating frames. In [5], the method by O'Malley et. al. [4] is modified using a textural descriptor for each frames and computes the dissimilarity matrix in a faster and robust way; moreover, the dynamic programming algorithm has been substituted by a local minima search in a 1D signal obtained from the dissimilarity matrix. It is important to stress that the quality of an image-based gating algorithm strongly depends on the ability of the method to extract a signal (1D or 2D) correlated with the actual vessel oscillation. In this paper, we present a novel and real-time image-based gating method that exploits *motion blur* variations during the cardiac cycle in coronary artery pullbacks.

Up to now, the experiments to validate image-gating techniques are questionable. The ECG signal used in the validation of [4] has been demonstrated to be a suboptimal descriptor of the relative oscillations between the catheter tip and the coronary artery [6]. In [5] the authors propose an indirect methodology to assess the smoothness of the reconstructed sequence.

In this paper a reliable validation of the technique is obtained by designing a novel ad-hoc in-vitro experiment in which ground-truth data was generated using an oscillating mechanical device. The performance of our algorithm is compared against the other state-of-the-art methods using such reference data.

2 Method

Image gating algorithms consist of two stages: the extraction of a 1D or 2D signal correlating with the cardiac cycle and the identification of the signal's instants corresponding to the most stable cardiac frames.

2.1 Cardiac Signal Extraction

The proposed cardiac signal extraction is based on the idea that, provided a high frame rate acquisition equipment, every tissue displacement causes, in the image being captured, a *motion blur* proportional to the speed of the tissue movement. We propose a *motion blur* intensity estimator $B(I)$ computed as minus the average of the absolute value of vertical derivative over the polar IVUS image:

$$B(I) = -E\left(\left|\frac{\delta I(r,\theta)}{\delta r}\right|\right) \tag{1}$$

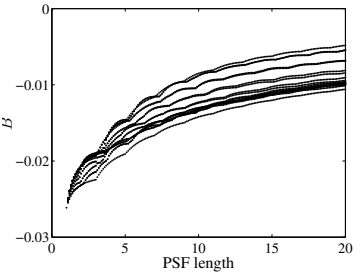


Fig. 1. Trend of the *motion blur* estimation on the simulated blurred IVUS images

where r is the radial position and θ is the angle of the probe during the image scan. The estimation is averaged (expected value E) over all the image A-lines in order to reduce outliers effect. In order to illustrate the suitability of the proposed estimator, we simulate increasing *motion blur* effect in a set of in-vivo IVUS image. Such phenomenon can be modeled by a convolution between the non-degraded image with a set of linear point spread functions (PSFs) with increasing length. This causes a blurring of the image, and thus the reduction of edge sharpness [7] [8]. Figure 1 shows the plot of the estimator B over the length of the simulated *motion blur* PSF. Since the actual PSF in clinical images is unknown, each PSF has been rotated in steps of 10°. This resulted in a set of curves exhibiting similar trends. The relationship between B and the PSF length is monotonically non decreasing and thus suitable to compute comparison between different frames. In order to obtain the variation of the *motion blur* during the sequence, we apply equation (1) to all the images in the sequence obtaining a signal $B(I_t)$, where t is the temporal variable, as depicted in Figure 2 (a). It is important to highlight that changes in the vessel structures, e.g. plaques, bifurcations, etc., modify the image texture thus influencing the signal $B(I_t)$, as it can be seen in Figure 2 (a). Since the pullback speed is significantly lower than the speed of the oscillation [4], we can separate the oscillatory part of the signal $B(I_t)$ from the changes in the vessel structure. This has been done by applying a Butterworth high-pass filter (order = 10) with cut-off at $0.5Hz$, obtaining the signal $B_{HP}(I_t)$. An oscillation below $0.5Hz$ is below 30 beat per minute (BPM) and thus it cannot be related to phenomena induced by heart beating. Figure 2 (b) shows the high-pass filtered version of the signal $B(I_t)$; periodic local minima, spaced about 1 second to the other, can be easily identified (see Figure 2 (c)). This repeating pattern has a frequency of $1.02 \pm 0.0\bar{6}$Hz which is in agreement with the ECG captured during the pullback ($1.13Hz$). The peak at about 1 Hz and the harmonics at multiple frequencies in Figure 2 (d) confirm the periodic nature of the signal. A possible explanation of the sharp periodical variations (from local minima to local maxima) in the $B_HP(I_t)$ signal can be drawn from the physiology of the cardiac cycle. The systolic phase is composed by the iso-volumetric contraction followed by the blood ejection from the left chamber. During the iso-volumetric contraction the pressure increases until the

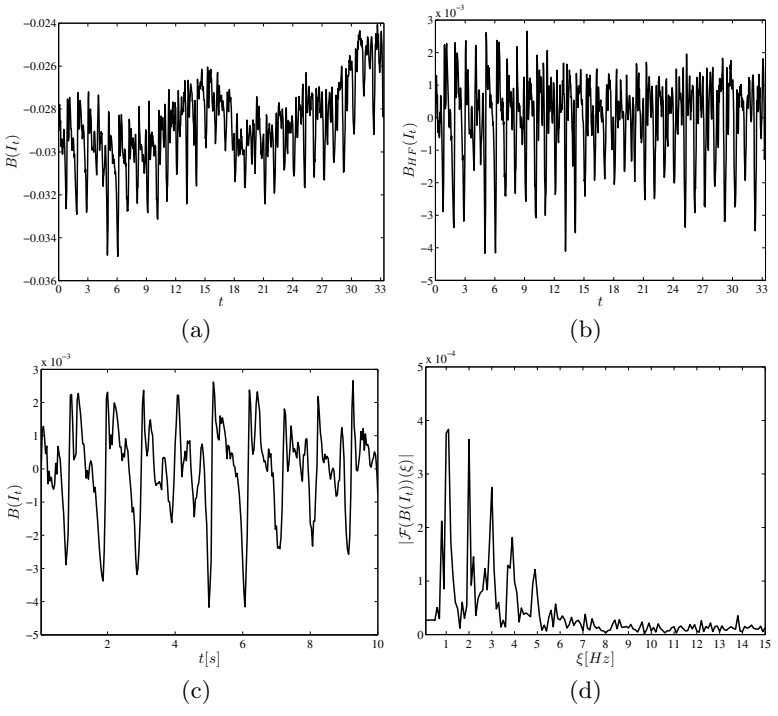


Fig. 2. *Motion blur* signal $B(I_t)$ obtained from a patient sequence (a); its respective high-pass filtered version $B_{HF}(I_t)$ (b); first 10 seconds of the signal $B_{HF}(I_t)$ (c); amplitude of the Fourier transform of $B_{HF}(I_t)$ (d)

ventricle pressure exceeds the aortic pressure, causing the abrupt opening of the aortic valve. This causes the sudden heart motion due to the rapid ejection of the blood, reflecting to a motion that can be captured by the proposed *motion blur* estimator.

2.2 Gating

Once the *motion blur* signal $B_{HF}(I_t)$ is computed, the stable frames in an arterial IVUS sequence must be reliably identified. Unfortunately not all the local minima corresponds to stable frames, so noise artifacts must be discarded. As it can be noticed in Figure 2 (c), the local minima corresponding to a stable frame are immediately followed by a local maximum. Exploiting this fact, we propose a method that selects a set of stable frame candidates among the local minima of $B_{HF}(I_t)$. A local minimum is retained as candidate if it is immediately followed by a local maximum which satisfy a zero-crossing criteria. The list of candidates is further refined by removing cases in which the gradient between the couple of minimum-maximum is lower than a certain threshold T. T is computed such that it separates the gradient distribution in two distinct modes.

3 Results

In this section, the proposed image-gating algorithm is quantitatively validated, and its performance are compared against the state-of-the-art methods. Reliable ground-truth data was obtained by designing a novel ad-hoc experiment.

3.1 *In vitro* Experiment

A mechanical device simulating the periodical heart movements has been designed in order to generate reliable ground-truth data. Figure 3 shows the mechanical device and its power supply. The apparatus is composed by a rotating eccentric wheel connected to a transversal arm generating horizontal swinging oscillations. The amplitude of the oscillation ΔL can be tuned using a regulating knob from 1 to 5 mm, in steps of 1 mm. The oscillation speed can be controlled setting the power supply voltage V, following the equation $F[BPM] = 4.1912\ [BPM/Volts]\ V\ [Volts]$. The catheter guide is connected to the oscillating arm. A total of 21 pullbacks from three *post-mortem* arteries have been acquired at 30 $frame/s$ using a Galaxy II IVUS Imaging System (Boston Scientific) and a catheter Atlantis SR Pro 40 MHz (Boston Scientific). For each artery 7 pullbacks have been recorded varying the mechanical device parameters ($\Delta L = \{1, 2, 3, 4\}$ mm, $F = 60$ [BPM] and $\Delta L = 3$ mm, $F = \{50, 60, 70, 80\}$ [BPM]). Each sequence contains 1000 frames, of 33.3 s duration. It is worth to note that the maximal frequency resolution using a 30 $frame/s$ acquisition setup is of $0.0\bar{6}$ Hz (4 BPM) and that the methodological incertitude induced by the power supply, (approximately ± 0.5 V), produces a methodological error in the oscillation frequency of ± 2 [BPM]. Such uncertainties must be accounted in the method evaluation. The sequences are available for research purpose and can be obtained under request.

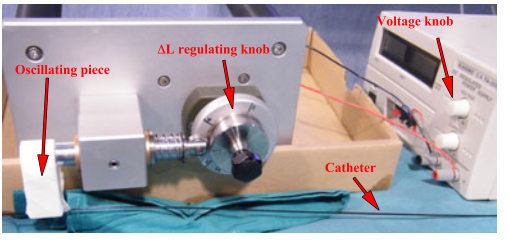


Fig. 3. Picture of the ad-hoc oscillating device and its power supply

3.2 Quantitative Evaluation

To quantitatively and fairly assess the performance of each image gating methods, we designed a measure that is robust to the methodological uncertainties.

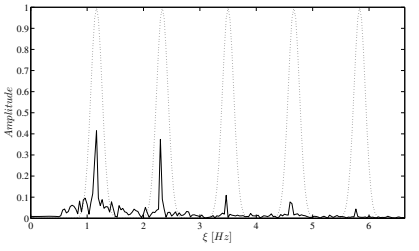


Fig. 4. An example of $|S(\xi)|$ and the filter $H_{f_N}(\xi)$, in solid and dashed lines respectively; in this example $f_N = 1.16$ Hz (70 BPM)

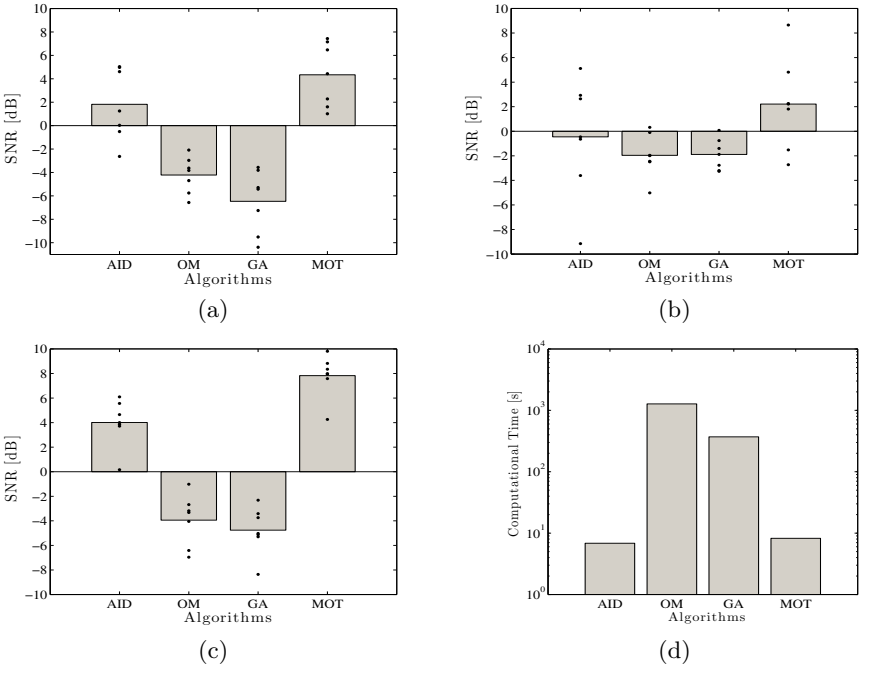


Fig. 5. SNR performance of different gating algorithms applied to three *post-mortem* arteries (a,b and c). Dots indicate the performance for each sequence while the gray bar is the average score. The plot in (d) shows the average computation (1000 frame sequence) for the 4 tested algorithms; it is important to note that the times scale is logarithmic.

For every sequence the nominal oscillation frequency f_N is known. Thus, we introduce the Signal to Noise Ratio (in dB) defined as follows:

$$SNR_{dB} = 20 \log \left(\frac{\int_\xi |S(\xi)| H_{f_N}(\xi)}{\int_\xi |S(\xi)| (1 - H_{f_N}(\xi))} \right) \tag{2}$$

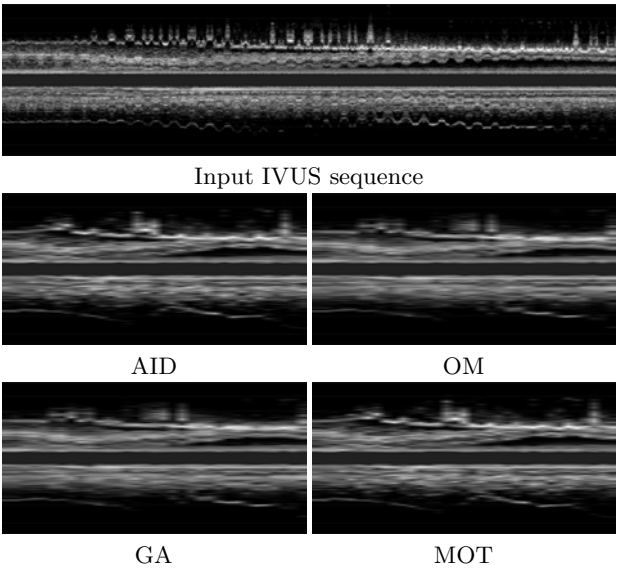


Fig. 6. Longitudinal cut of an input IVUS sequence and the reconstructed longitudinal cuts obtained using different image-based gating algorithms

where $S(\xi)$ is the Fourier transform of the signal $s(t)$ produced by a given algorithm; H is a frequency filter designed to filter out the frequencies that are not related with the nominal one (f_N). Hence, the numerator of equation (2) represents the energy of the signal, related to the nominal frequency f_N, while the denominator corresponds to the energy of the signal not related with the oscillation, i.e. the noise of the signal. In such way, the higher is the SNR, the better the image-gating technique extracts the oscillatory component from the image sequence. The filter $H(\xi)$ is designed as follows:

$$H_{f_N}(\xi) = \sum_{n=1}^{M} \exp\left(-\frac{(\xi - nf_N)^2}{2\sigma_M^2}\right) \tag{3}$$

An example of $|S(\xi)|$ and the filter $H_{f_N}(\xi)$, are shown in figure 4 in solid and dashed lines respectively. Since the catheter can bend inside the artery, the catheter tip oscillation is not perfectly sinusoidal. To account for this mechanical limitation, the filter of equation (3) includes higher harmonics. Using the defined SNR measure, we tested the methods in [3] (AID method), [4] (OM algorithm) and [5] (GA) compared with our technique (MOT). Figure 5 (a-c) shows the results on three different post-mortem arteries. As it can be notice, the proposed method outperforms other methods in all three arteries. However, for the artery in figure 5 (b), the dispersion of AID and MOT is large. The ANOVA test confirmed that the results of our method are statistically significant with respect to AID for the first two arteries, while it is not for the third one. Figure 5 (d) shows the average computational time, for a 1000 frame sequence, in

seconds. The computational time of our method is of the same order of the fastest algorithm (AID) as it can be seen in figure 5 (d).

3.3 Qualitative In-vitro Evaluation

A qualitative comparison of the image-gating methods is illustrated in figure 6, showing a longitudinal cut of an IVUS sequence and the resulting longitudinal cuts obtained using the four gating algorithms. The input longitudinal cut has been obtained from a sequence of the proposed ground-truth data; its appearance is very similar to in-vivo acquired sequences, showing the typical saw-tooth shaped artifact. It can be noticed that the result of OM and GA are very similar since GA is an improvement of OM. In this sequence, the difference between AID and MOT is not easily noticeable, while their superior performance with respect to OM and GA is evident, especially observing the arterial tissue sharpness.

4 Conclusion

In this paper we proposed a real-time gating algorithm based on the analysis of *motion blur* variations during the IVUS sequence. For the first time in literature, a reliable validation of image-based gating techniques is performed. The algorithm have been tested on ground-truth data generated by an ad-hoc oscillating mechanical device. The proposed method outperforms the other state-of-the-art algorithms both in SNR and computational complexity. Future work will be addressed to an extensive validation of the method with several post-mortem arteries.

Acknowledgments

This work was supported in part by the research projects TIN2009-14404-C02-02, FIS PI061290 y CSD2007-00018. The work of C. Gatta and S. Balocco is supported by the Catalan Agency for Administration of University and Research (AGAUR) under a Beatriu de Pinos Fellowship.

References

1. Arbab-Zadeh, A., DeMaria, A.N., Penny, W., Russo, R., Kimura, B., Bhargava, V.: Axial movement of the intravascular ultrasound probe during the cardiac cycle: Implications for three-dimensional reconstruction and measurements of coronary dimensions. American Heart Journal 138, 865–872 (1999)
2. Bruining, N., von Birgelen, C., de Feyter, P.J., Ligthart, J., Li, W., Serruys, P.W., Roelandt, J.: Ecg-ated versus nongated three-dimensional intracoronary ultrasound analysis: Implications for volumetric measurements. Catheterization and Cardiovascular Diagnosis 43, 254–260 (1998)
3. Zhu, H., Oakeson, K.D., Friedman, M.H.: Retrieval of cardiac phase from IVUS sequences. In: Medical Imaging 2003: Ultrasonic Imaging and Signal Processing, vol. 5035, pp. 135–146 (2003)

4. O'Malley, S.M., Granada, J.F., Carlier, S., Naghavi, M., Kakadiaris, I.A.: Image-based gating of intravascular ultrasound pullback sequences. IEEE Transactions on Information Technology in Biomedicine 12(3), 299–306 (2008)
5. Gatta, C., Pujol, O., Leor, O.R., Ferre, J.M., Radeva, P.: Robust image-based ivus pullbacks gating. Med. Image Comput. Assist. Interv. Int. 11(Pt 2), 518–525 (2008)
6. O'Malley, S.M., Carlier, S.G., Naghavi, M., Kakadiaris, I.A.: Image-based frame gating of IVUS pullbacks: A surrogate for ecg. In: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007, April 15-20, vol. 1, pp. I-433–I-436 (2007)
7. Yitzhaky, Y., Kopeika, N.S.: Identification of blur parameters from motion blurred images. Graphical Models and Image Processing 59(5), 310–320 (1997)
8. Dai, S., Wu, Y.: Motion from blur. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008, pp. 1–8 (June 2008)

Registration of a Statistical Shape Model of the Lumbar Spine to 3D Ultrasound Images

Siavash Khallaghi[1], Parvin Mousavi[1], Ren Hui Gong[1], Sean Gill[1], Jonathan Boisvert[2], Gabor Fichtinger[1], David Pichora[3], Dan Borschneck[3], and Purang Abolmaesumi[4]

[1] Queen's University, Kingston, ON, Canada
[2] National Research Council, Ottawa, ON, Canada
[3] Kingston General Hospital, ON, Canada
[4] University of British Columbia, Vancouver, BC, Canada
purang@ece.ubc.ca

Abstract. *Motivation:* Spinal needle injections are technically demanding procedures. The use of ultrasound image guidance without prior CT and MR imagery promises to improve the efficacy and safety of these procedures in an affordable manner. *Methodology:* We propose to create a statistical shape model of the lumbar spine and warp this atlas to patient-specific ultrasound images during the needle placement procedure. From CT image volumes of 35 patients, statistical shape model of the L3 vertebra is built, including mean shape and main modes of variation. This shape model is registered to the ultrasound data by simultaneously optimizing the parameters of the model and its relative pose. Ground-truth data was established by printing 3D anatomical models of 3 patients using a rapid prototyping. CT and ultrasound data of these models were registered using fiducial markers. *Results:* Pairwise registration of the statistical shape model and 3D ultrasound images led to a mean target registration error of 3.4 mm, while 81% of all cases yielded clinically acceptable accuracy below the 3.5 mm threshold.

1 Introduction

Spinal needle injection is widely applied in analgetic and diagnostic purposes [1]. In the management of back pain, epidural anesthesia, facet joint injections and nerve blocks are common forms of these interventions, performed in great numbers in hospitals and radiology clinics. Back pain is the second most common reason for a visit to the physician. In the United States alone, approximately 90% of adults will experience back pain at some point in their life. Nearly 50% of the current working population has reported some history of back pain; this is the leading cause for missed work time and seriously degrades on-the-job performance [2]. In back pain management, accessing the lumbar epidural space presents major difficulties. The success rate of proper needle insertion is a dismal 60% after 10 attempts [3] when no guidance is used, clearly an inadequate practice. Contemporary radiological needle guidance with CT and fluoroscopy

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 68–75, 2010.

requires specialized facilities often unavailable to patients living in rural areas, and involve X-ray radiation. In search of a more accessible, portable, and non-toxic imaging alternative, ultrasound guidance has been considered. Watson et al. [4] and Klocke et al. [5] targeted the L3-L4 and L2-L3 interspace, a challenging procedure that, if performed inaccurately, damages the spinal cord. The results suggested that ultrasound as a solo guidance is inadequate. To address this issue, combination of ultrasound with CT has been proposed [6,7,8]. In these studies, the auxiliary information is necessary for guidance, but as stated above, either unobtainable or involves ionizing radiation. Hence, the use of statistical shape models is a logical alternative.

Previously in the literature, statistical shape models (or atlases) have been considered as an alternative to pre-operative CT scans [9,10,11]. These atlases are generally divided into two main categories: those that describe the outline of objects (geometrical atlases) [9,10,11,12,13] and those that contain both the geometrical and internal density distribution of the object (volumetric atlases) [14,15,16]. As an example of geometrical atlases, [17] captures the variations in pose of each vertebra and builds a model for the curvature of the spine. The feasibility of ultrasound registration to a statistical atlas of femur has been previously investigated [18]. In [12] a geometrical atlas to ultrasound registration of the pelvis is performed. While geometrical atlases are computationally less expensive, they are prone to bone surface segmentation errors in ultrasound data.

Our contribution is the first report of a volumetric vertebral atlas and its registration to 3D ultrasound without prior segmentation. We demonstrate a feasibility study on the L3 vertebrae, one of the most problematic anatomical sites in spinal pain management. In departure from the prior art using tetrahedral mesh and Bernstein polynomials [13], we employ a simple but generic approach based on the B-spline deformable transformation that allows for straightforward extension for an ensemble of vertebrae and relevant structures [14,15,16].

2 Method

2.1 Construction of the Statistical Shape Model

Figure 1 demonstrates the atlas model construction process. A set of CT images, acquired from 38 patients (19 male and 19 female), was used in this study. Data was collected under the approval of the research ethics board, and the patients involved provided informed consent for this data to be used in the research. Using ITK-Snap, the L3 vertebra was semi-automatically segmented from the CT volumes (each containing $120 \times 200 \times 100$ voxels with an isotropic spacing of 0.6 mm). The patient data was divided into two groups: 35 for constructing the atlas (hereafter referred to as training data), and 3 for validation (two male and one female). One of the CT volumes for atlas construction was chosen as the template I_{t}. Each training example, I_k, is registered to the template by a rigid registration followed by a B-spline deformable registration, such that $I_{\mathrm{t}} \approx T^k_{\mathrm{def}}(T^k_{\mathrm{rigid}}(I_k))$, where *T(.)* denotes a transform. B-spline registration is performed in a $40 \times 30 \times 30$ grid using Mattes Mutual Information metric. To

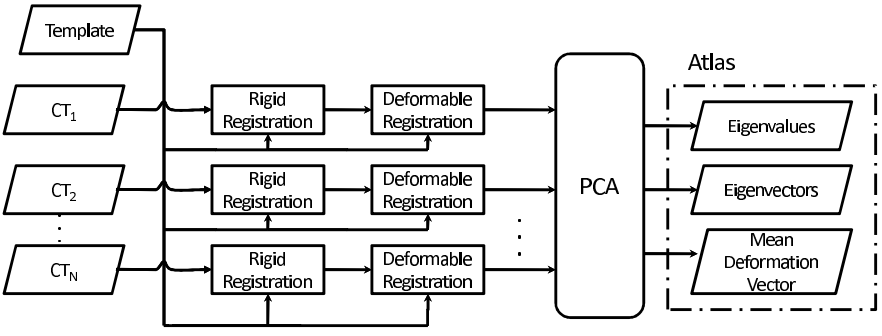


Fig. 1. Outline of the atlas construction method from a set of CT images

reduce the deformable registration time, a three stage multi-resolution approach is implemented. Each deformable registration took between two and four hours on a Core 2 Quad CPU machine with 2.4 GHz speed and 3 GB of RAM. With the deformable transform of all the training examples known with respect to the template, principal component analysis (PCA) is performed to construct the statistical atlas for the L3 vertebrae.

After the atlas is constructed it can be used to generate new instances of the population. A new instance of the atlas, defined by the deformation vector, D_{new} ,can be produced by a linear combination of the mean deformation vector, $\bar{\phi}$, atlas weights, w_i, and the eigenvectors of the covariance matrix generated from all the deformation fields, v_i, as follows:

$$D_{new} = \bar{\phi} + \sum_{i=1}^{N} w_i v_i \tag{1}$$

2.2 Statistical Atlas to Ultrasound Registration

The registration framework is shown in Figure 2. First, the mean shape of the L3 atlas is rigidly registered to the 3D ultrasound volume by simulating ultrasound images from the atlas, and performing an intra-modality registration with the 3D ultrasound images [19]. The similarity metric used is the Linear Correlation of Linear Combinations (LC^2) [19] between the actual and the simulated ultrasound images, and is computed from the atlas volume using:

$$LC^2 = \frac{\sum (U(x,y) - f(x,y))^2}{M \times Var(U)} \tag{2}$$

where f is the simulated ultrasound image, U is the actual ultrasound image and M is the number of overlapping voxels between the ultrasound and the atlas volume. We used the Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

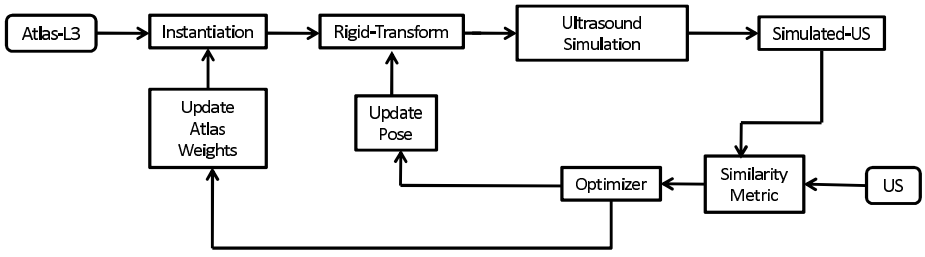


Fig. 2. Outline of statistical atlas to ultrasound registration method

as the optimization method [20]. Following the initial rigid registration, we performed deformable registration of the atlas model to the 3D ultrasound data by simultaneously creating new instances of the atlas and updating the rigid transformation and the atlas parameters, while optimizing the LC^2 similarity metric.

3 Experiments and Results

3.1 Statistical Shape Model

We investigated whether the statistical shape model generated from L3 can span the space of shape variations in the patient population. For this purpose, we used a leave-one-out cross registration with the CT data. In each step, we left out one of the patient CTs for testing, created the atlas from the rest of the data, and registered the created atlas to the remaining patient CT. This process was repeated by selecting each CT data once as the testing data. Throughout the process, the template was held constant. To validate the model, we used the first 12 eigenvectors which covered 95% of the total modes of variation. The registration process used Mattes Mutual Information and deformable B-Spline registration, in conjunction with the CMA-ES optimizer to solve for the atlas model and rigid registration parameters.

As in [9], the Root Mean Square (RMS) error between the closest surface points of registered atlas and the test data was used to measure the capability of the atlas to capture the deformations within the patient population. The mean error across all registrations was 0.89 mm with a standard deviation of 0.19 mm.

3.2 Registration of the Atlas to Ultrasound Images

In a second set of experiments, we aim to show how well the constructed atlas can be deformed and reoriented to match the shape and position of the patient vertebra using the acquired 3D ultrasound images. In each registration, we simultaneously optimized 12 shape and 6 rigid parameters, which correspond to the first 12 PCA eigenvectors and the pose respectively.

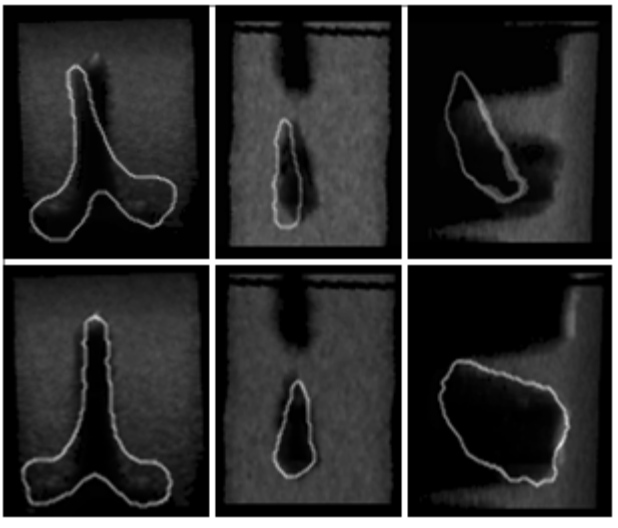
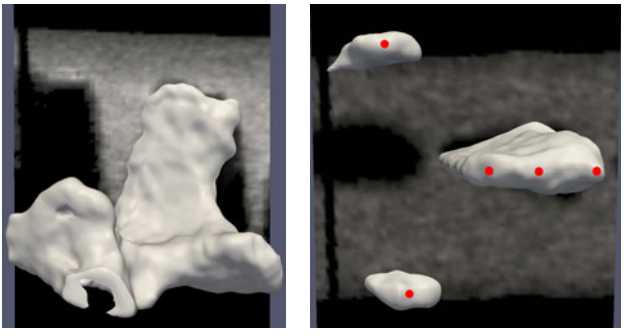


Fig. 3. Transverse (left), sagittal (center), and coronal (right) slices of the original US volume overlayed with the bone contours of the misaligned (top) and registered (bottom) atlas volumes

The three excluded CT volumes from the atlas generation process (see Section 2.1) were used to construct 3D CAD models of the entire lumbar spine, including L1 to L5. These models were printed using a Cimetrix 3D shape printer (Cimetrix Solutions, Oshawa, ON, Canada). Three spine phantoms were constructed by submerging these models in an agargelatine-based tissue phantom which was designed to simulate the appearance of soft tissue in ultrasound. A high-resolution CT image ($0.46 \times 0.46 \times 0.625$ mm) and an ultrasound volume were acquired from each phantom. The ultrasound volume was reconstructed from a freehand sweep with an L14-5/38 linear-array transducer (Ultrasonix, Richmond, BC, Canada) operating at 6.6 MHz with an imaging 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 phantom [21]. The phantom CT and ultrasound volumes were aligned using fiducial markers mounted on the exterior of the phantom box. The position of these fiducials in the ultrasound coordinate system was identified using a calibrated stylus pointer.

The atlas mean shape and the ultrasound volumes were brought to an initial position by rigidly registering the mean shape to the corresponding phantom CT volume. For each phantom, thirty experiments were performed with perturbing the mean shape using a transformation generated from a uniform random distribution in the interval of [0,10] mm translation along each axis and [0-10]° rotation about each axis. The registration parameters were then optimized as it was discussed in Section 2.2.



(a) (b)

Fig. 4. (a) Sagittal, and (b) coronal views of the registered L3 atlas to the 3D ultrasound volume. The five red circles illustrate the landmark positions.

To evaluate the accuracy of the registration, two expert orthopedic surgeons were asked to identify five corresponding landmarks, three on the spinous process and two on the target facet joints shown in Figure 4(b), on the registered atlas, the ultrasound volume and the corresponding CT. The average distance of these five landmarks was chosen as a measure of the final Target Registration Error (TRE). A registration was considered failed if the final TRE is more than 3.5 mm, as the clinically accepted error. Registration results are shown in Table 1 and an example of the initial misalignment and the registration result is depicted in Figure 3. Figure 4 shows an overlay of the registered atlas with the ultrasound volume.

As seen in Table 1, the average TRE is less than 3.5 mm with success rate of 81% for all three phantoms. Our preliminary results for the registration of atlases created from L2 and L4, the neighboring vertebrae to L3, to 3D ultrasound volumes, with the same patient data set, have shown TRE below 3.2 mm based on five landmarks per vertebrae. The registration results satisfy clinical requirements for facet joint injection. They also demonstrate the feasibility of using volumetric atlases for the registration of a patient spine to 3D ultrasound data. This can enable the use of ultrasound image guidance for spinal interventions without prior CT, to improve the efficacy of these procedures in an affordable manner.

On going research is aimed at addressing several goals: i) Run time; currently, a single registration of the atlas model to ultrasound data, implemented in unoptimized C++ code, takes in the order of hours on a 2.3 GHz, 16-core Pentium machine with 16 GB of RAM. Future efforts will focus on speeding up the computation, specifically by implementing the algorithm on graphics processing units (GPUs). Recently released open-source software for GPU implementation of B-spline interpolation and registration will facilitate achieving this goal; ii) Atlas generation bias; the current atlas generation process is potentially biased towards the chosen template. An alternative would be to use groupwise atlas generation methods that have become widely popular recently, especially in neuroimaging

Table 1. Registration results for the atlas to US registration. SR (Success Rate) is defined as the ratio of the registrations where the overall TRE is less than 3.5 mm. SR is presented for each phantom with the maximum initial misalignment of 10 mm.

Phantom	Mean Landmark Error mm	Std mm	SR %
1	3.38	0.42	81%
2	3.48	0.33	79%
3	3.25	0.45	82%

research; iii) Spine registration; at the moment, the method only registers the atlas of a single vertebra to 3D ultrasound volumes. Registration of an atlas of the entire (or partial) spine with the ultrasound data would provide better contextual information to the physician for intervention.

References

1. An, H.S.: Percutaneous procedures in the lumbar spine. In: Principles and Techniques of Spine Surgery. Lippincott Williams & Wilkins (1998)
2. Allen, H., Hubbard, D., Sullivan, S.: The burden of pain on employee health and productivity at a major provider of business services. Journal of Occupational and Environmental Medicine 47(7), 658–670 (2005)
3. Watts, R.: A five-year prospective analysis of the efficacy, safety and morbidity of epidural anaesthesia performed by a general practitioner anaesthetist in an isolated rural hospital. Anaesth Intensive Care 20(3), 348–353 (1992)
4. Watson, M.J., Evans, S., Thorp, J.M.: Could ultrasonography be used by an anaesthetist to identify a specified lumbar interspace before spinal anaesthesia? British Journal of Anaesthesia 90(4), 509–511 (2003)
5. Klocke, R., Jenkinson, T., Glew, D.: Sonographically guided caudal epidural steroid injections. Journal of Ultrasound in Medicine 22(11), 1229–1232 (2003)
6. Moore, J., Clarke, C., Bainbridge, D., Wedlake, C., Wiles, A., Pace, D., Peters, T.: Image Guidance for Spinal Facet Injections Using Tracked Ultrasound. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 516–523. Springer, Heidelberg (2009)
7. 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. IEEE Transactions on Evolutionary Computation 12(3), 284–296 (2008)
8. Gill, S., Mousavi, P., Fichtinger, G., Chen, E., Boisvert, J., Pichora, D., Abolmaesumi, P.: Biomechanically Constrained Groupwise US to CT Registration of the Lumbar Spine. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 803–810. Springer, Heidelberg (2009)
9. Tang, T.S., Ellis, R.E.: 2D/3D Deformable Registration Using a Hybrid Atlas. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 223–230. Springer, Heidelberg (2005)

10. Hu, Y., Ahmed, H.U., Allen, C., Pendse, D., Sahu, M., Emberton, M., Hawkes, D., Barratt, D.: MR to Ultrasound Image Registration for Guiding Prostate Biopsy and Interventions. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 787–794. Springer, Heidelberg (2009)
11. Michopoulo, S.K., Costaridou, L., Panagiotopulous, E., Speller, R., Panayiotakis, G., Todd-Pokropek, A.: Atlas-Based Segmentation of Degenerated Lumbar Interverbal Discs from MR Images of the Spine. IEEE Transactions on Biomedical Engineering 56(9), 2225–2231 (2009)
12. Foroughi, P., Song, D., Chintapani, G., Taylor, R.H., Fichtinger, G.: Localization of Pelvic Anatomical Coordinate System Using US/Atlas Registration for Total Hip Replacement. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 871–878. Springer, Heidelberg (2008)
13. Sadowsky, O., Cohen, J.D., Taylor, R.H.: Rendering Tetrahedral Meshes with Higher-Order Attenuation Functions for Digital Radiograph Reconstruction. In: IEEE Visualization, pp. 303–310 (2005)
14. Barratt, D.C., Chan, C.S.K., Edwards, P.J., Penney, G.P., Slomczykowski, M., Carter, T.J., Hawkes, D.J.: Instantiation and Registration of Statistical Shape Models of the Femur and Pelvis Using 3D Ultrasound Imaging. Medical Image Analysis 12(3), 358–374 (2008)
15. Gong, R.H., Stewart, J., Abolmaesumi, P.: Reduction of Multi-Fragment Fractures of the Distal Radius Using Atlas-based 2D/3D Registration. In: Proceedings of the SPIE, Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling, vol. 7261, pp. 726137-1–726137-9 (2009)
16. Jurcak, V., Fripp, J., Engstrom, C., Walker, D., Salvado, O., Ourselin, S., Crozier, S.: Atlas Based Automated Segmentation of the Quadratus Lumborum Muscle Using Non-Rigid Registration on Magnetic Resonance Images of the Thoracolumbar Region. In: 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, pp. 113–116 (2008)
17. Boisvert, J., Cheriet, F., Pennec, X., Labelle, H., Ayache, N.: Geometric Variability of the Scoliotic Spine Using Statistics on Articulated Shape Models. IEEE Transactions on Biomedical Engineering 27(4), 557–568 (2008)
18. Talib, H., Rajamani, K., Kowal, J., Nolte, L.P., Styner, M., Ballester, M.A.G.: A Comparison Study Assessing the Fesibility of Ultrasound-Initialized Deformable Bone Models. Computer Aided Surgery 10(5/6), 293–299 (2005)
19. 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)
20. Hansen, N., Müller, S.D., Koumoutsakos, P.: Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation. Evolutionary Computation 11(1), 1–18 (2003)
21. Chen, T.K., Thurstonlow, A.D., Ellis, R.E., Abolmaesumi, P.: A real-time freehand ultrasound calibration system with automatic accuracy feedback and control. Ultrasound in Medicine & Biology 1(1), 79–93 (2009)

Automatic Prostate Segmentation Using Fused Ultrasound B-Mode and Elastography Images

S. Sara Mahdavi[1], Mehdi Moradi[1], William J. Morris[2], and Septimiu E. Salcudean[1]

[1] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada
{saram,moradi,tims}@ece.ubc.ca
[2] Vancouver Cancer Center, British Columbia Cancer Agency, Vancouver, Canada
jmorris@bccancer.bc.ca

Abstract. In this paper we propose a fully automatic 2D prostate segmentation algorithm using fused ultrasound (US) and elastography images. We show that the addition of information from mechanical tissue properties acquired from elastography to acoustic information from B-mode ultrasound, can improve segmentation results. Gray level edge similarity and edge continuity in both US and elastography images deform an Active Shape Model. Comparison of automatic and manual contours on 107 transverse images of the prostate show a mean absolute error of 2.6 ± 0.9 mm and a running time of 17.9 ± 12.2 s. These results show that the combination of the high contrast elastography images with the more detailed but low contrast US images can lead to very promising results for developing an automatic 3D segmentation algorithm.

1 Introduction

Low dose rate (LDR) prostate brachytherapy is a common method for treating patients with low risk prostate cancer. In this treatment, 40-100 small radioactive seeds are permanently inserted in the prostate and its periphery. Treatment planning and delivery relies on transrectal ultrasound (TRUS) imaging. In order to create the treatment plan, a pre-operative volume study is carried out in which a set of transverse ultrasound images are collected. These images are then manually segmented to extract the prostate boundaries. A plan is devised to deliver sufficiently high radiation dose to the cancerous tissue while maintaining a tolerable dose to healthy tissue. Reliable segmentation and visualization of the prostate is a vital step in dose planning. Manual segmentation is time consuming and, due to the low signal to noise ratio of ultrasound images, inter and intra-observer variabilities are high. Even though various 2D prostate segmentation methods and some 3D methods have been proposed in the literature [1,2,3,4,5], the effective automatic segmentation of ultrasound images of the prostate has remaining challenges such as user initialization and limited accuracy.

In this paper, we propose a prostate segmentation method based on intra-modality fusion of ultrasound B-mode and elastography. Elastography [6], in

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 76–83, 2010.

which mechanical properties of tissue are characterized, has shown to be promising in improving the visibility of the prostate gland [7,8]. In a recent study, we have shown that ultrasound dynamic elastography images of the prostate have superior object-background contrast compared to B-mode ultrasound, especially at the base and apex [9]. This is due to the fact that prostate tissue is generally stiffer than the surrounding tissue. We utilize this advantage and combine elastography and US image data for 2D segmentation of the prostate. The elastography images are acquired using a system described in [9] which enables the simultaneous registered acquisition of B-mode and elastography images, thereby eliminating the concerns about image registration and cost.

We use an Active Shape Model (ASM) [10] approach which starts with an initial shape extracted from a large number of elastography and B-mode prostate images. The deformation of this initial shape is restricted to conform to the statistical model and is guided by edge detection from both elastography and B-mode images based on edge gradient similarity and continuity. The use of a statistically created model ensures the compliance of the resulting contours with the overall shape of the organ. Additionally, the restricted deformation results in robustness to poor image quality. The use of a measure of edge continuity [9] in addition to gradient similarity, reduces the effects of strong speckle-induced local edges on the algorithm which improves the rate of convergence. We provide a statistical analysis of the accuracy of our 2D image segmentation method and show that the combined use of elastography and B-mode images improves the accuracy and the convergence rate. Further, we describe a preliminary framework for extending the proposed method to an automatic 3D segmentation algorithm and present an example.

2 Methods

The US and elastography images used in this paper were acquired from patients going through the standard LDR prostate brachytherapy procedure at Vancouver Cancer Center, BC Cancer Agency. Intra-operatively, prior to the procedure, RF data and US images were simultaneously collected using the system described in [9]. In this system RF data is collected at approximately 40 fps from the sagittal array of a vibrating (amplitude 0.5-2 mm, frequency range 2-10 Hz) and rotating ($-45°$ to $50°$) TRUS probe (dual-plane linear/microconvex broadband 5 - 9 MHz endorectal transducer, Ultrasonix Medical Corp.). The RF data were processed [11] to obtain sagittal elastography strain images from which conventional transverse images of the prostate were achieved.

Our 2D segmentation approach combines the information from elastography and US images within an Active Shape Model (ASM) which deforms based on gray level similarity and edge continuity. We follow the approach from [10]. First, we construct a training set by manually selecting 30 specific points on the prostate boundary in $N = 25$ mid-gland images from 7 patients. The set of images used for creating the training set does not include the images to be segmented. The manually segmented contours are aligned by using least-squares-based iterative scaling, rotation and translation and the average of the resulting

contours, $\bar{x}$, is used as the initial contour. Then, in order to capture the statistics of the training set, we calculate the covariance matrix, $S = \sum_{i=1}^{N} dx_i dx_i^T$, where dx_i in the training phase is the distance between each point on the manual contour and the corresponding point on the mean shape. The modes of variation of the shape are described by the eigenvalues, λ_i, and eigenvectors, p_i, of S, from which the t largest eigenvalues are selected as the most significant modes of variation. We selected t such that for $i = 1, \cdots, t$, $\lambda_i / \sum \lambda > 5\%$. A shape instance consistent with the training set can thus be created using Eq.(1):

$$x = \bar{x} + Pb \tag{1}$$

where $P = (p_1 \cdots p_t)$ is the matrix of the first t eigenvectors and $b = (b_1 \cdots b_t)^T$ is a vector of weights. Hence, in the shape fitting phase, given a shape deformation, dx, the shape parameter adjustments, db, can be calculated using Eq.(2).

$$db = P^T dx \tag{2}$$

To calculate the movement, dx_i, for each model point i, a measure of edge gray level similarity was used in [10]. For each point i of each image j of the training set, a normalized edge derivative profile, g_{ij} normal to the boundary, centered at the model point and of length n_p is extracted. g_{ij} is averaged over all images from which the covariance matrix S_{g_i} is calculated to obtain a statistical description of the gray level appearance for every point. During the shape fitting, at each iteration and for every point, sample edge derivative profiles $h_i(d)$ of distance d $(d = -l, \cdots, l)$ from the boundary point and length n_p are extracted in a similar manner. The square of the Mahalanobis distance of these profiles from the model profile, g_i, give a measure of edge gray level similarity. For each point, the d resulting in the least Mahalanobis distance suggests the required point movement along a line normal to the boundary. The physical values of l and n_p are set to 8 and 5 mm, respectively.

In our data set, we observed that the gray level edge similarity measure alone, gives many false positives due to ultrasound speckle or sharp edge-like structures since only 1D information (normal to the edge) is being analyzed. Therefore, in our approach, we incorporate our edge continuity measure [9] which measures the continuity of the edge in a direction orthogonal to the edge profile. At a distance d from each point, we compute the average normalized cross-correlation, $c_i(d)$, between the edge intensity profile $e_i(d)$ (obtained similar to g_{ij}), of length n_p, with its two neighboring left and right edge intensity profiles. For a continuous edge (i.e. large similarity between $e_i(d)$ and its left and right neighboring profiles), $c_i(d)$ should have a shape similar to a Gaussian function with a large peak and a small standard deviation. We define the edge continuity measure, $K_i(d) = P_{c_i}^2 / \sigma_i(d)$ in which P_{c_i} characterizes the peak and σ_i is the standard deviation of a Gaussian function fitted to $c_i(d)$. For each point, the d resulting in the maximum continuity measure suggests the required point movement along a line normal to the boundary.

We define u_n to be the vector normal to the boundary at the boundary point, d_{gE}, to be the distance from the boundary point suggested by the gray level

similarity, and d_{KUS}, and d_{KE} to be the distances from the boundary point computed from the edge continuity measures in US and elastography images, respectively. Based on the above equations, at each iteration, k, the following steps are performed to deform the current shape points, $x_{current}$ into the next shape, x_{next}.

1. Find the required shape deformation:
   $dx = x_{next} - x_{current} = d_f u_n$
   $d_f = \alpha_1 d_{gE} + \alpha_2 d_{KUS} + \alpha_3 d_{KE}$
2. Calculate the optimum pose parameters: scaling, translation and rotation, corresponding to dx, apply this transformation to obtain $T(x_{current})$. Adjusting the pose is required to align $x_{current}$ to be as close as possible to x_{next} before adjusting the shape [10].
3. Calculate $db = P^T(x_{next} - T(x_{current}))$

$\alpha = [\alpha_1 \quad \alpha_2 \quad \alpha_3]$ are corresponding weights. Due to the large amount of noise in US images, gray level similarity matching in these images does not improve results but degrades convergence, and therefore, it is not included. Our criteria for convergence is when 94% of the contour points have a dx of less than $n_p/2$. Fig. 1 illustrates how d_{gE} and d_{KE} are obtained in an elastography image.

We will provide a comparison of segmentation results using gray level similarity from elastography images, combined edge continuity and gray level similarity from elastography images and finally edge continuity and gray level similarity from both elastography and US images.

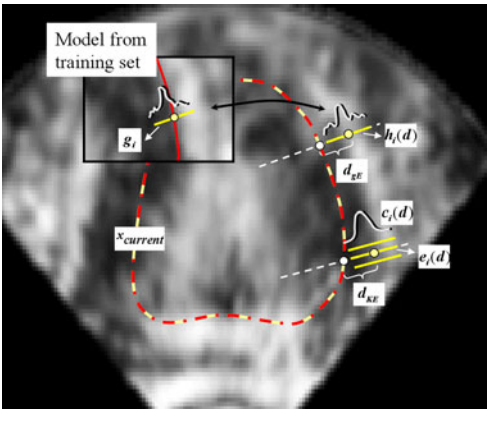


Fig. 1. An illustration of how d_{gE} (top point) and d_{KE} (bottom point) are obtained in an elastography image. The gray level similarity measure is compared to that of the model for each corresponding point. The edge continuity measure is maximized over the line normal to the boundary. For clarity, these measures are shown on two different points, whereas they are calculated for single points.

3 Results

To evaluate our 2D segmentation results we measure the mean absolute distance (MAD) and maximum distance (MaxD) between 2D automatic and manual contours. Table 1 provides the mean and standard deviations of MAD and MaxD between 107 manual and automatic contours selected from 7 patients. The results are presented separately for elastography gray level similarity only (α_a=[1 0 0]), elastography gray level similarity and edge continuity (α_b=[0.5 0 0.5]), and elastography gray level similarity and edge continuity plus US edge continuity (α_c=[0.5 0.25 0.25]). Fig. 2 shows an example of segmentation results for the three sets of weight parameters.

The most accurate segmentation results were acquired when α_c=[0.5 0.25 0.25]. By using this selection of weight parameters, deformation is mainly guided by the coarser elastography images but also refined by the finer US images. It is specifically seen in the posterior region of the prostate, where elastography image quality is low, that the addition of edge continuity in US images improves segmentation results. This can be observed in Fig. 2. The choice of α_c also results in the convergence of the algorithm in an average of 22 iterations vs. 37 and 98 iterations for α_b and α_a. The maximum number of iterations was set to 50 for α_b and α_c and 100 for α_a. In the case of α_a, 95% of the cases did not converge within 100 iterations.

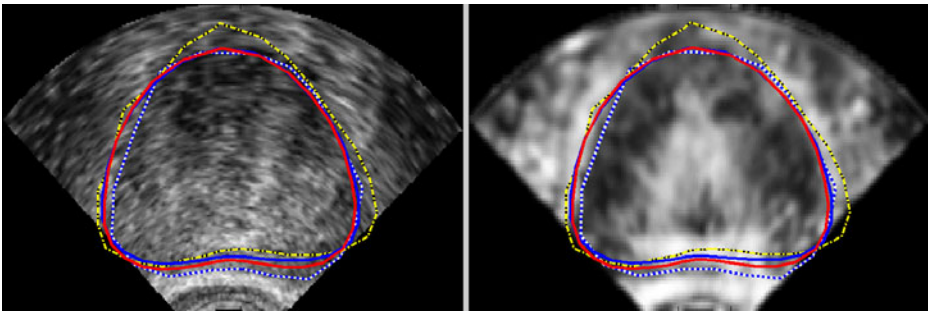


Fig. 2. 2D automatic segmentation results using elastography gray level similarity (yellow dashed line), elastography edge continuity and gray level similarity (blue line), and elastography gray level similarity and edge continuity plus US edge continuity (red line) on US (left) and elastography (right) images. The manual contour is shown as a blue/white dotted line.

4 Discussion

By visual inspection of the automatic segmentation results, we observed that most of the 2D segmentation error of the mid-gland slice was in the anterior and posterior regions. In elastography images, the blood vessels which lie on the anterior of the prostate appear as stiff tissue and are not always distinguishable

Table 1. Comparison of 2D manual and automatic segmentation showing the Mean Absolute distance (MAD) and Maximum Distance (MaxD - positive sign meaning larger automatic contour) between manual and automatic prostate contours, the number of iterations and duration of the algorithm. K: edge continuity measure, d_g: gray level similarity measure.

	d_g **in elast.**	d_g **and K in elast.**	d_g **and K in elast. and K in US**
MAD (mm)	3.4 ± 1.8	3.4 ± 1.7	2.6 ± 0.9
MaxD (mm)	1.0 ± 9.8	-6.5 ± 7.0	-4.9 ± 4.7
no. of iter.	98 ± 13	37 ± 17	22 ± 15
duration (s)	10.7 ± 1.4	17.3 ± 0.8	17.9 ± 12.2

from the prostate itself. In such cases, the automatic contour extends beyond the actual boundary. In the posterior, due to the relatively low contrast in this region, the automatic contour converges to the darker tissue inside the prostate. By including edge continuity data from US images this problem has been partially resolved. We attribute this low contrast to mainly the slippage between the protective sheath on the probe and the surface of the rectum during elastography data collection. By replacing the 1D axial strain computation with 2D axial/lateral and by increasing the resolution of elastography imaging such problems can be subsided.

Our current elastography data acquisition system has the benefit of collecting inherently registered elastography and US data. Currently the TRUS rotation range is within $\pm 50^\circ$ which may result in missing data in the mid-lateral regions of large prostates. Also, the quality of the B-mode US images acquired along with the elastography data is affected by the computational limitations of the real-time data acquisition system and the ultrasound machine. We are currently working on resolving these problems to improve segmentation results.

The proposed 2D prostate segmentation method using fused elastography and US image information can be extended into 3D by modifying the method that we proposed in [12] which was based on fitting an *a priori* shape to a set of parallel transverse ultrasound images. The algorithm was initialized by the user manually selecting some initial boundary points on the mid-slice. These points were used to un-warp and un-taper all images resulting in a set of elliptical prostate shapes. With the aid of the Interacting Multiple Model Probabilistic Data Association (IMMPDA) [13] edge detector and ellipse fitting, a tapered ellipsoid was fitted to all contours which was then sliced at image depths. The resulting 2D contours were reversely tapered and warped to match the initial images. We showed [12] that the method is fast, and produces smooth contours that are in agreement with the brachytherapy requirements. However, the need for manual initialization limits its use for real-time applications and makes it user dependent. Also, the poor visibility of the boundary at the base and apex complicates segmentation of these regions.

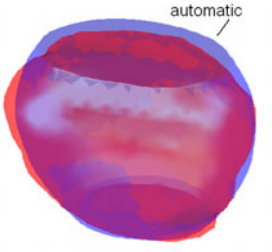


Fig. 3. An example of 3D surface of the prostate created manually (red) and by the automatic algorithm (blue)

To resolve these issues we replace the manual initialization of the 3D semi-automatic segmentation with the described 2D segmentation. Additionally, we employ elastography along with US images for propagating the mid-gland segmentation to the rest of the images. This is done by including an additional IMMPDA edge detection on the coarser elastography images to guide the edge detection on the finer US images. The described framework for automatic 3D prostate segmentation was applied to one patient data set. Fig. 3 shows 3D surfaces created from automatic (blue) and manual (red) segmentation of the prostate for this patient. 11 images were used to construct this surface model. The volume of the manually created surface is 33.9 ml, the volume of the automatically created surface is 34.4 ml, and the volume of the non-overlapping region between the two surfaces is 4.9 ml. A thorough clinical study is required to evaluate this 3D segmentation framework.

5 Conclusions

In this paper we outlined a novel 2D method of prostate segmentation that combines ultrasound elastography imaging with B-mode data. This is the first instance of using such a combination for prostate segmentation and reinforces efforts to improve US segmentation outcomes using elastography data [14]. With the fusion of information from elastic properties of tissue provided by elastography with the acoustic properties of tissue provided by B-mode we developed an automatic and accurate segmentation of the prostate which gives good results in 2D. The automatically generated 2D contours can be used to initialize the mid-slice for 3D segmentation and remove user variability. Additionally, the method can be utilized to register pre and intra-operative prostate images and has the potential of improving intra-operative dosimetry.

Acknowledgments. This project was funded by NSERC Canada (Discovery Grant and Postdoctoral Fellowship). Data collection was funded by NIH grant R21 CA120232-01.

References

1. Penna, M.A., Dines, K.A., Seip, R., Carlson, R.F., Sanghvi, N.T.: Modeling prostate anatomy from multiple view TRUS images for image-guided HIFU therapy. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 54(1), 52–69 (2007)
2. Tutar, I.B., Pathak, S.D., Gong, L., Cho, P.S., Wallner, K., Kim, Y.: Semiautomatic 3-D prostate segmentation from TRUS images using spherical harmonics. IEEE Trans. Med. Imaging 25(12), 1645–1654 (2006)
3. Zhan, Y., Shen, D.: Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method. IEEE Trans. Med. Imaging 25(3), 256–272 (2006)
4. Hodge, A.C., Fenster, A., Downey, D.B., Ladak, H.M.: Prostate boundary segmentation from ultrasound images using 2D active shape models: optimization and extension to 3D. Comput. Methods Programs Biomed. 84(2-3), 99–113 (2006)
5. Shen, D., Zhan, Y., Davatzikos, C.: Segmentation of prostate boundaries from ultrasound images using statistical shape model. IEEE Trans. Med. Imaging 22(4), 539–551 (2003)
6. Ophir, J., Cépedes, I., Ponnekanti, H., Yazdi, Y., Li, X.: Elastography: a quantitative method for imaging the elasticity of biological tissues. Ultrason. Imaging 13(2), 111–134 (1991)
7. Cochlin, D.L., Ganatra, R.H., Griffiths, D.F.R.: Elastography in the detection of prostatic cancer. Clin. Radiol. 57(11), 1014–1020 (2002)
8. Souchon, R., Hervieu, V., Gelet, A., Ophir, J., Chapelon, J.: Human prostate elastography: in vitro study. In: IEEE Symposium on Ultrasonics, vol. 2, pp. 1251–1253 (2003)
9. Mahdavi, S.S., Moradi, M., Wen, X., Morris, W.J., Salcudean, S.E.: Vibro-elastography for visualization of the prostate region: method evaluation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 339–347. Springer, Heidelberg (2009)
10. Cootes, T., Taylor, C., Cooper, D., Graham, J.: Active shape models-their training and application. Comput. Vis. Image Underst. 61, 38–59 (1995)
11. Zahiri-Azar, R., Salcudean, S.E.: Motion estimation in ultrasound images using time domain cross correlation with prior estimates. IEEE Trans. Biomed. Eng. 53(10), 1990–2000 (2006)
12. Mahdavi, S.S., Morris, W.J., Spadinger, I., Chng, N., Goksel, O., Salcudean, S.E.: 3D prostate segmentation in ultrasound images based on tapered and deformed ellipsoids. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 960–967. Springer, Heidelberg (2009)
13. Abolmaesumi, P., Sirouspour, M.R.: An interacting multiple model probabilistic data association filter for cavity boundary extraction from ultrasound images. IEEE Trans. Med. Imaging 23(6), 772–784 (2004)
14. von Lavante, E., Noble, J.A.: Segmentation of breast cancer masses in ultrasound using radio-frequency signal derived parameters and strain estimates. In: IEEE ISBI, pp. 536–539 (2008)

ODF Maxima Extraction in Spherical Harmonic Representation via Analytical Search Space Reduction

Iman Aganj[1], Christophe Lenglet[1,2], and Guillermo Sapiro[1]

[1] Department of Electrical and Computer Engineering, University of Minnesota, USA
[2] Center for Magnetic Resonance Research, University of Minnesota, USA
{iman,clenglet,guille}@umn.edu

Abstract. By revealing complex fiber structure through the orientation distribution function (ODF), q-ball imaging has recently become a popular reconstruction technique in diffusion-weighted MRI. In this paper, we propose an analytical dimension reduction approach to ODF maxima extraction. We show that by expressing the ODF, or any antipodally symmetric spherical function, in the common fourth order real and symmetric spherical harmonic basis, the maxima of the two-dimensional ODF lie on an analytically derived one-dimensional space, from which we can detect the ODF maxima. This method reduces the computational complexity of the maxima detection, without compromising the accuracy. We demonstrate the performance of our technique on both artificial and human brain data.

1 Introduction

Diffusion-weighted MRI significantly extends the scope of the information obtained from MRI, from being solely spatially dependent to being defined on the spatial-orientational domain. Fiber microstructure and orientation are inferred using this modality from the locally measured diffusion profile of water molecules. Diffusion tensor imaging (DTI) [1] effectively models the diffusion in single-fiber voxels as a Gaussian represented by its covariance tensor. As for more complex fiber architecture, q-ball imaging (QBI) [2]–[6] has been very successful in revealing intravoxel fiber orientations by introducing the orientation distribution function (ODF) as the probability of diffusion in a given direction.

Contrary to DTI, where the principal diffusion direction can be readily computed as the major eigenvector of the diffusion tensor, QBI provides a continuous spherical function which, although clearly illustrates the major diffusion orientations as its maxima, does not directly quantify them. Diffusion directions as vectors carry less information than the ODF itself does. On the other hand, their easy interpretation and their application in tractography, e.g., [7]–[9], make the ODF maxima extraction an important post-processing step still to be carefully addressed. The number of peaks can also be interpreted as a measure of white matter complexity. In addition, unlike mixture models that calculate fiber directions by describing the diffusion signal as the sum of finite discrete unidirectional components, ODF maxima are computed without any assumptions about the existence of such components.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 84–91, 2010.

Exhaustive search via finite difference method has been exploited in the literature as a straightforward approach to ODF maxima extraction [3,10]. This generally requires a two-dimensional (2D) discretization of the unit sphere, resulting in computational complexity that grows quadratically with the desired resolution. Numerical optimization approaches such as gradient ascent [11], Newton-Raphson techniques [12], and Powell's method [13], have also been employed. These techniques require a guarantee of convergence and careful initialization to obtain all the maxima. Lastly, polynomial based approaches, [14]–[16], have been proposed to extract the maxima as a subset of the stationary points of the ODF. These methods exploit a transformation of the real and symmetric spherical harmonic (RSSH) basis (most efficient for ODF reconstruction [3]), to the constrained symmetric tensor or constrained homogenous polynomial bases, resulting in polynomial equations which are solved numerically.

In this paper, we propose a polynomial based approach to reduce the problem of ODF maxima extraction in the fourth order RSSH basis, from a 2D search on the sphere, to a one-dimensional (1D) one on an analytically-derived curve. Compared to the 2D problem, this approach significantly reduces the computational complexity of the search for the maxima of the ODF – or any antipodally symmetric spherical function – without compromising the precision. Contrary to [14]–[16], our method works directly in the RSSH basis and does not require the extra step of transforming the RSSH coefficients to other tensor-based bases. We suggest a discretization scheme for the 1D exhaustive search, and show experimental results on both artificial and human brain data.

We start Sec. 2 with a brief review of the RSSH basis, and continue by describing our mathematical derivation. Experimental results are presented in Sec. 3.

2 Methods

2.1 ODF in Real and Symmetric Spherical Harmonic Basis

In this work, we use the estimator derived in [6] to compute the ODF in constant solid angle (CSA). The original definition of the QBI ODF [2] does not include the Jacobian factor r^2, thereby creating the need for normalization and artificial sharpening. In contrast, the estimator in [6] is normalized, dimensionless, and has been shown to preserve the natural sharpness of the ODF.

The spherical harmonic basis is commonly used for representing spherical functions such as the ODF, allowing for sampling in any desired direction. Orthonormal spherical harmonic functions are given by

$$Y_l^m(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\phi}, \tag{1}$$

where $P_l^m(\cdot)$ is the associated Legendre function, and θ and ϕ are standard spherical coordinates. The assumption of the ODF being real and antipodally symmetric,

however, makes the use of the RSSH basis [3] more suitable. RSSH functions are indexed by a single parameter $j = l(l+1)/2 + m + 1$, corresponding to l_j and m_j, as follows [3]:

$$Y_j = \begin{cases} (-1)^{m_j}\sqrt{2}\mathrm{Re}\left\{Y_{l_j}^{-m_j}\right\}, & -l_j \le m_j < 0 \\ Y_{l_j}^{0}, & m_j = 0 \\ \sqrt{2}\mathrm{Im}\left\{Y_{l_j}^{m_j}\right\}, & 0 < m_j \le l_j. \end{cases} \tag{2}$$

The ODF can be computed in this basis first by using a minimum square scheme to approximate the signal, and then by analytically computing the Funk-Radon transform, [2], following the method introduced in [3]–[5] for the original QBI, and subsequently adapted in [6] for the CSA-QBI.

2.2 ODF Maxima Extraction

RSSH functions, being smooth, allow us to find all the local maxima of the ODF $\psi(\theta, \phi)$ as points satisfying the following properties (subscripts indicate partial derivatives):

$$\psi_\theta(\theta, \phi) = 0, \tag{3}$$
$$\psi_\phi(\theta, \phi) = 0, \tag{4}$$
$$\det\big(H(\theta, \phi)\big) \ge 0, \tag{5}$$
$$\mathrm{tr}\big(H(\theta, \phi)\big) \le 0, \tag{6}$$

with the Hessian matrix $H(\theta, \phi)$ defined as

$$H(\theta, \phi) = \begin{pmatrix} \psi_{\theta\theta}(\theta, \phi) & \psi_{\theta\phi}(\theta, \phi) \\ \psi_{\theta\phi}(\theta, \phi) & \psi_{\phi\phi}(\theta, \phi) \end{pmatrix}. \tag{7}$$

Equations (3) and (4) guarantee that (θ, ϕ) is either an extremum or a saddle point of the ODF. Inequalities (5) and (6) filter out, respectively, the saddle points and the local minima (including possible negative lobes), leaving us only with the local maxima of the ODF. The above expressions can all be analytically computed for an ODF expressed in the RSSH basis. However, the main challenge is to find the points that simultaneously satisfy equations (3) and (4). Once they are identified, applying inequalities (5) and (6) to filter out undesired points is trivial.

Iterative approaches (e.g., Newton method) may be applied to solve equations (3) and (4). Yet, being quite sensitive to the initialization, they are not guaranteed to converge to all the maximum points. Alternatively, an exhaustive search will result in all the maxima with an accuracy determined by the discretization resolution. Nonetheless, with the ODF being a 2D manifold, the search space, and consequently the computational complexity of the algorithm, grows quadratically with the desired resolution.

We will next show how the fourth order RSSH basis makes it possible to confine the search to a 1D space, thereby creating an efficient method to extract the maxima.

2.3 Reducing the Dimension of the Search Space

Let us assume that the ODF has been approximated in the fourth order RSSH basis, as

$$\psi(\theta,\phi)=\sum_{j=1}^{15} a_j Y_j(\theta,\phi). \tag{8}$$

Combining equations (1), (2), and (8), while substituting the values of $P_l^m(\cos\theta)$ from Table 1 leads to

$$\begin{aligned}\psi(\theta,\phi)=\frac{a_1}{2\sqrt{\pi}}&+A(3\cos^2\theta-1)+B(\phi)\sin\theta\cos\theta+C(\phi)\sin^2\theta\\&+D(35\cos^4\theta-30\cos^2\theta+3)\\&+E(\phi)(7\cos^2\theta-3)\sin\theta\cos\theta\\&+F(\phi)(7\cos^2\theta-1)\sin^2\theta+G(\phi)\cos\theta\sin^3\theta\\&+H(\phi)\sin^4\theta,\end{aligned} \tag{9}$$

where $A=\sqrt{\frac{5}{16\pi}}a_4$, $B(\phi)=\sqrt{\frac{15}{8\pi}}(a_3\cos\phi-a_5\sin\phi)$, etc. (We drop the notation (ϕ) in the rest of this subsection.)

We now attempt to solve Eq. (4) by deriving Eq. (9) with respect to ϕ. We then divide it by $\sin\theta\cos^3\theta$ and rearrange it, while using the identity $\sec^2\theta=1+\tan^2\theta$, to obtain

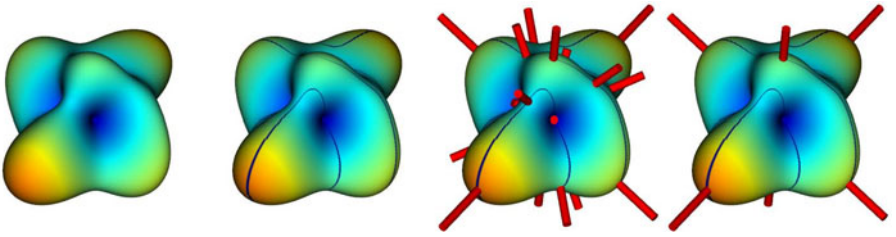
$$\begin{aligned}(H+C-F)_\phi\tan^3\theta+(G+B-3E)_\phi\tan^2\theta+(6F+C)_\phi\tan\theta\\+(B+4E)_\phi=0.\end{aligned} \tag{10}$$

Equation (10) is a cubic function of $\tan\theta$, and can be analytically solved, leading to a closed-form expression for $\theta(\phi)$.[1] Thus, for each given ϕ, we obtain one, two, or three different real values for θ which satisfy Eq. (4).

The curve characterized by the pair $(\theta(\phi),\phi)$ (Fig. 1(b)) is in fact our new 1D search space which contains all the ODF maxima as points satisfying equations (3), (5), and (6) (Fig. 1(c&d)). The number of these maxima does not need to be initially specified, since it is automatically determined by the algorithm and depends on various factors, such as the number of real solutions to Eq. (10). This is particularly important in practice, as different regions of the white matter naturally exhibit different complexity. The maxima can be found using a 1D exhaustive search (see Sec. 2.4), which is considerably faster than exploring the entire 2D manifold of the ODF.[2]

[1] Each solution of $\tan\theta$ corresponds to a unique value of $\theta\in[0,\pi)$. Please note that this approach can also be applied in the RSSH basis of higher orders, with the difference that there may be no analytical solution for $\theta(\phi)$, and numerical methods may need to be applied.

[2] Such 1D exhaustive searches can also be performed using tensor-based approaches [14]–[16].



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

Fig. 1. (a) Reconstructed ODF. (b) Analytically defined 1D space is searched. (c) All the extrema and saddle points are identified. (d) ODF maxima are extracted.

Table 1. The associated Legendre functions required for the proposed algorithm

Function		Expression
$P_0^0(\cos\theta)$	=	1
$P_2^0(\cos\theta)$	=	$(1/2)(3\cos^2\theta - 1)$
$P_2^1(\cos\theta)$	=	$-3\cos\theta\sin\theta$
$P_2^2(\cos\theta)$	=	$3\sin^2\theta$
$P_4^0(\cos\theta)$	=	$(1/8)(35\cos^4\theta - 30\cos^2\theta + 3)$
$P_4^1(\cos\theta)$	=	$-(5/2)(7\cos^2\theta - 3)\cos\theta\sin\theta$
$P_4^2(\cos\theta)$	=	$(15/2)(7\cos^2\theta - 1)\sin^2\theta$
$P_4^3(\cos\theta)$	=	$-105\cos\theta\sin^3\theta$
$P_4^4(\cos\theta)$	=	$105\sin^4\theta$

2.4 One-Dimensional Exhaustive Search

Here we detail the discretization scheme used to perform the aforementioned 1D exhaustive search for the maxima. We exploit the closed-form description of the curve $(\theta(\phi), \phi)$ provided by Eq. (10) and parameterize the curve with $\phi \in [0,2\pi)$. To achieve a constant spatial resolution $\Delta s = \sqrt{\Delta\theta^2 + \sin^2\theta\,\Delta\phi^2}$, we need a variable step size $\Delta\phi$:

$$\Delta\phi = \frac{\Delta s}{\sqrt{\theta'^2(\phi) + \sin^2\theta(\phi)}} = \frac{1 + t^2(\phi)}{\sqrt{t'^2(\phi) + t^2(\phi) + t^4(\phi)}}\Delta s, \tag{11}$$

which is rewritten as a function of $t(\phi) \coloneqq \tan\theta(\phi)$. For every ϕ, Eq. (10) results in one, two or three real values for $t(\phi)$, for each of which $t'(\phi)$ can be computed simply by deriving Eq. (10) with respect to ϕ, and substituting for ϕ and t. Therefore, at each step we choose $\Delta\phi$ to be the minimum of the three (or fewer) values obtained from Eq. (11).

Next, we keep all the *candidate* points satisfying inequalities (5), (6), and the following, which is a relaxation of Eq. (3),

$$|\psi_\theta(\theta, \phi)| < \alpha. \tag{12}$$

We found an appropriate value of $\alpha = 0.02\sim0.03$ for the threshold. Note again that inequalities (5), (6), and (12) can all be computed analytically using equations (1), (2), (8), and Table 1. The ODF maxima are then computed as the mean directions

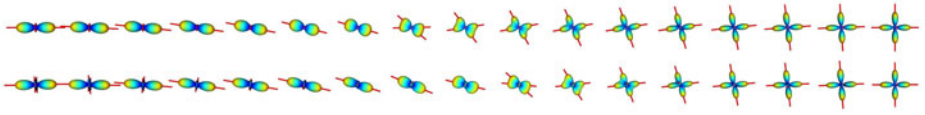


Fig. 2. Extracted maxima from synthetic ODFs with fiber crossing, in noise-free case (top), and with SNR=40 (bottom)

corresponding to the *clusters* of points, created by processing all the candidate points, as follows: Each point is added to a previous cluster if its Euclidean distance to the representative (mean) point of that cluster is minimum among all other clusters and is smaller than a threshold (0.4 was used here). If no such cluster is found, a new cluster is created, and the algorithm goes on until all the candidate points are processed.

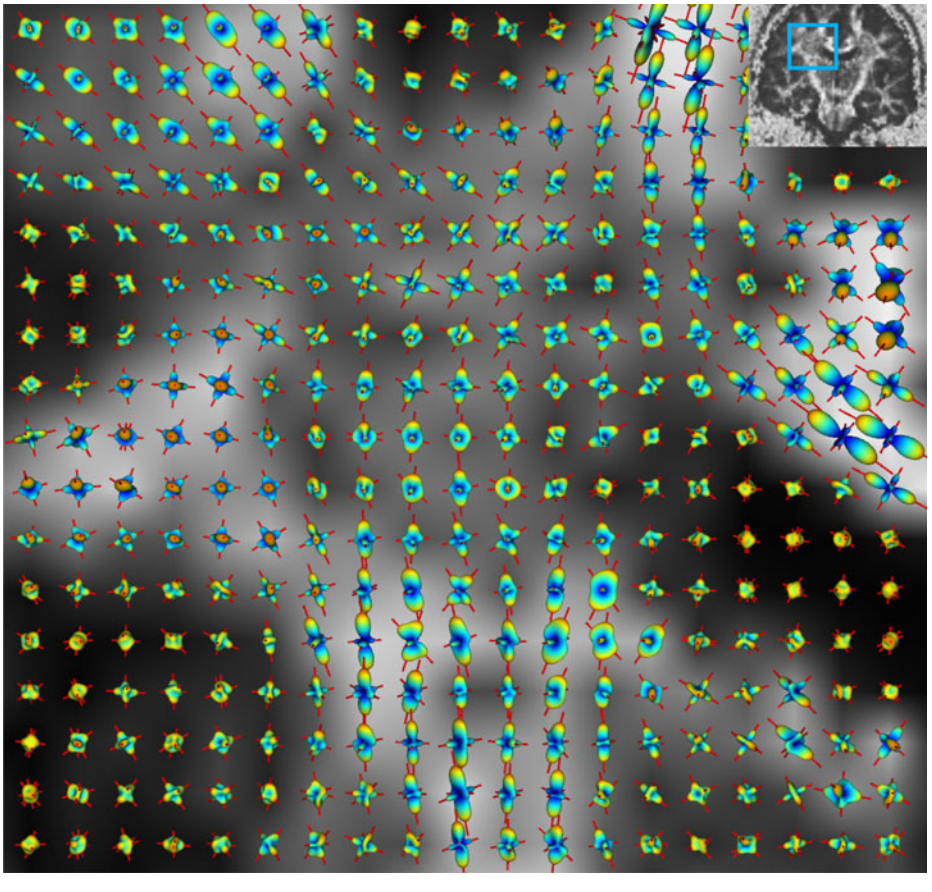


Fig. 3. Experimental results on human brain data, superimposed on the FA map

3 Results and Discussion

To validate our approach, we first show results on artificial data. We simulated fiber crossing by generating diffusion images from the sum of two exponentials, $S(\hat{u}) = \left(e^{-\hat{u}^T D_1 \hat{u}} + e^{-\hat{u}^T D_2 \hat{u}}\right)/2$, where D_1 is a diagonal matrix with diagonal entries (9, 2, 2), and D_2 is D_1 rotated about the z-axis by a varying angle. CSA-ODFs were reconstructed in the fourth order RSSH basis from 76 diffusion directions, uniformly sampled on the sphere. The maxima were then extracted using the proposed technique, and results are depicted in Fig. 2 (top). Increasing the angular precision to 0.5° revealed that multiple fiber orientations are resolved starting at the crossing angle of 37.5°. Choosing a spatial resolution of $\Delta s = 0.001$, required the evaluation of the ODF at 7.7×10^4 points, whereas a 2D search on the sphere with the same resolution would cost 1.6×10^7 operations. When we repeated the experiment by adding Rician noise with a signal-to-noise ratio (SNR) of 40 (Fig. 2, bottom), the minimum angle where crossing was detected increased to 48°. Such experiments are commonly employed to evaluate the robustness of the ODF reconstruction algorithm to noise.

We also tested our method on a popular public human brain dataset [17]. CSA-ODFs were reconstructed in the fourth order RSSH basis from 200 diffusion images acquired at b=3000 s/mm². Figure 3 illustrates the ODFs with their extracted maxima superimposed on the fractional anisotropy (FA) map, in the region of the centrum semiovale, where three major fiber bundles intersect. To demonstrate the performance of the proposed technique, *all* the maxima are shown here, including those corresponding to slight variations in the ODF (for example due to noise). Major ODF peaks corresponding to fiber orientations may however be selected by placing a threshold on the ODF [3] or on its curvature [15].

Acknowledgments. This work was partly supported by NIH (P41 RR008079, P30 NS057091, R01 EB007813, R01 MH060662, R01 EB008432, CON000000004051–3014), NSF, ONR, NGA, ARO, and DARPA. We would like to thank the authors of [17] for providing us with human brain data.

References

1. Basser, P., Mattiello, J., LeBihan, D.: Estimation of the effective self-diffusion tensor from the NMR spin echo. J. Magn. Reson. B. 103(3), 247–254 (1994)
2. Tuch, D.: Q-ball imaging. Magnetic Resonance in Medicine 52(6), 1358–1372 (2004)
3. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical q-ball imaging. Magnetic Resonance in Medicine 58(2), 497–510 (2007)
4. Anderson, A.: Regularized, fast, and robust analytical q-ball imaging. Magnetic Resonance in Medicine 54(5), 1194–1206 (2005)
5. Hess, C., Mukherjee, P., Han, E., Xu, D., Vigneron, D.: Q-ball reconstruction of multimodal fiber orientations using the spherical harmonic basis. Magnetic Resonance in Medicine 56(1), 104–117 (2006)
6. Aganj, I., Lenglet, C., Sapiro, G., Yacoub, E., Ugurbil, K., Harel, N.: Reconstruction of the orientation distribution function in single and multiple shell q-ball imaging within constant solid angle. Magnetic Resonance in Medicine 64(2), 554–566 (2010)

7. Mori, S., Crain, B., Chacko, V., Van Zijl, P.: Three dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Annals of Neurology 45(2), 265–269 (1999)
8. Conturo, T., Lori, N., Cull, T., Akbudak, E., Akbudak, A., Shimony, J., McKinstry, R., Burton, H., Raichle, M.: Tracking neuronal fiber pathways in the living human brain. Proc. of National Academy of Sciences 96(18), 10422–10427 (1999)
9. Behrens, T., Johansen Berg, H., Jbabdi, S., Rushworth, M., Woolrich, M.: Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? NeuroImage 34(1), 144–155 (2007)
10. Frey, S., Campbell, J., Pike, G., Siddiqi, K.: Dissociating the human language pathways with high angular resolution diffusion fiber tractography. Journal of Neuroscince 28(45), 11435–11444 (2008)
11. Berman, J., Chung, S., Mukherjee, P., Hess, C., Han, E., Henry, R.: Probabilistic streamline q-ball tractography using the residual bootstrap. NeuroImage 39(1), 215–222 (2008)
12. Tournier, J., Calamante, F., Gadian, D., Connelly, A.: Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. NeuroImage 23(3), 1176–1185 (2004)
13. Jansons, K., Alexander, D.: Persistent angular structure: new insights from diffusion magnetic resonance imaging data. Inverse Problems 19, 1031–1046 (2003)
14. Ghosh, A., Tsigaridas, E., Descoteaux, M., Comon, P., Mourrain, B., Deriche, R.: A polynomial based approach to extract the maxima of an antipodally symmetric spherical function and its application to extract fiber directions from the Orientation Distribution Function in Diffusion MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242. Springer, Heidelberg (2008)
15. Bloy, L., Verma, R.: On computing the underlying fiber directions from the diffusion orientation distribution function. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 1–8. Springer, Heidelberg (2008)
16. Qi, L., Han, D., Wu, E.: Principal invariants and inherent parameters of diffusion kurtosis tensors. Journal of Mathematical Analysis and Applications 349(1), 165–180 (2009)
17. Poupon, C., Poupon, F., Allirol, L., Mangin, J.: A database dedicated to anatomofunctional study of human brain connectivity. In: Proc. of the 12th Annual Meeting of OHBM (2006)

An Anthropomorphic Polyvinyl Alcohol Triple-Modality Brain Phantom Based on Colin27

Sean Jy-Shyang Chen[1], Pierre Hellier[2], Jean-Yves Gauvrit[2,3,4], Maud Marchal[2], Xavier Morandi[2,3,4], and D. Louis Collins[1]

[1] McConnell Brain Imaging Centre, Montreal Neurological Institute, McGill University, Montreal, Canada
sjschen@bic.mni.mcgill.ca
[2] INRIA, Centre de Recherche, Rennes, Bretagne Atlantique, France
[3] INSERM U746, IRISA, F-35042, Rennes, France
[4] Université de Rennes 1, 35065, Rennes, France

Abstract. We propose a method for the creation of an anatomically and mechanically realistic brain phantom from polyvinyl alcohol cryogel (PVA-C) for validation of image processing methods for segmentation, reconstruction, registration, and denoising. PVA-C is material widely used in medical imaging phantoms for its mechanical similarities to soft tissues. The phantom was cast in a mold designed using the left hemiphere of the Colin27 brain dataset [1] and contains deep sulci, a complete insular region, and an anatomically accurate left ventricle. Marker spheres and inflatable catheters were also implanted to enable good registration and simulate tissue deformation, respectively. The phantom was designed for triple modality imaging, giving good contrast images in computed tomography, ultrasound, and magnetic resonance imaging. Multimodal data acquired from this phantom are made freely available to the image processing community (http://pvabrain.inria.fr) and will aid in the validation and further development of medical image processing techniques.

1 Introduction

The human cerebrum is a topologically complex organ with deep fissures and sulci over its lateral and medial surfaces, as well as fluid filled ventricles of complex form in its interior. The creation of a physical model capable of depicting the form of the cerebrum in a realistic manner is not trivial due in part to these deep structures. Previous works in creating brain phantoms have either reduced the depth of the sulci [2], or only recreated the brain's form superficially with dessert gelatin molds [3,4]. Although these phantoms bear a gross cursory resemblance to the human cerebrum, they do not accurately depict its anatomy. Registering these phantoms to their acquired multi-modality images may also not be straight-forward since the landmarks on the phantom are not easy to find or image. This may be due to the structures being smaller than the image resolution or having insufficient contrast of the markers with respect to the surrounding

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 92–100, 2010.

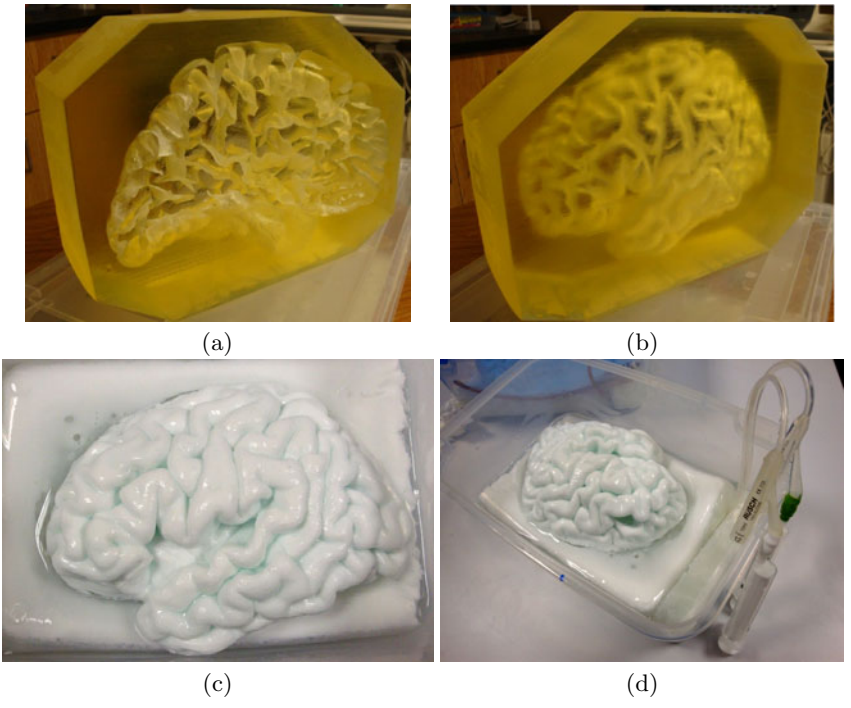


Fig. 1. Two views of the elastic Colin27 based brain phantom mold from (a) the inside and (b) the outside. (c) The PVA-C phantom casted from our mold using our PVA solution recipe. Note the deep sulci and insular regions of the phantom. (d) The setup for scanning the phantom with the catheters used to inflate the phantom on the right.

tissues. For instance, Reinertsen and Collins [4] rely on the presence of bubbles in their phantom to act as landmarks for validation.

Our goal is to create a triple modality human brain phantom containing anatomically realistic structures and physically realistic texture. The phantom was then scanned and the multimodal images are made publically available. Polyvinyl alcohol was selected for phantom construction.

Polyvinyl alcohol (PVA) is a synthetic polymer synthesized from polyvinyl acetate through hydrolysis of the latter's acetate groups [2]. When liquid PVA solutions undergo a specified period of freezing at a set temperature and is then allowed to slowly thaw to room temperature, this freeze-thaw cycle (FTC) transforms the liquid PVA solution into a elastic semi-opaque gel know as polyvinyl alcohol cryogel (PVA-C) [5,6,7,8].

PVA-C is used in biomedical research for producing soft tissue phantoms in studies to develop, characterize, and refine different imaging or image processing methods [9,10,11]. It is a good material for such studies since it has similar texture, mechanical properties such as compressibility and elasticity, and similar water content to many soft tissues[2,4,9,10,11,12,13]. PVA-C has been used in

the construction of phantoms for studying a wide variety of tissues including that of the heart [14], breast [10,11], prostate [13], arterial vasculature [8,2], the brain [4,2], as well as their abnormal tissues in the form of lesions and tumours [10,11,13].

In the following sections, we describe the methods we used to build our brain phantom. The contributions of our phantom to the literature include:

Anatomical accuracy: Deep cortical structures of the Colin27 cerebrum, such as the sulci, the insular region and the ventricles are realistically represented in the cast phantom.

Realistic texture: Recipes of PVA-C with textures most similar to human cerebral tissues were determined through the feedback of an experienced neurosurgeon who knows the tactility of the human brain and tumour tissues.

Multimodal imaging: PVA-C formulations that can be imaged effectively with high contrast CT, US, and MR are used to construct the phantom and its implants.

Freely available data: Images acquired using the US, MR, and CT scanners are made available through our website to researchers and the general public. Images for this phantom were acquired using magnetic resonance imaging (T1 and T2 weighted, PD, FLAIR, and DTI), ultrasound imaging, and computed tomography, to ensure that the phantom exhibits similar contrast to images of the live cerebrum acquired using these imaging modalities.

2 Phantom Construction

2.1 PVA Solution Preparation

The PVA solutions used to cast the brain phantom and its various components were prepared using 99-100% hydrolyzed PVA with an average molecular weight of 86 kilodaltons (Code:418120010) from Acros organics (Geel, Belgium).

Master solutions of 5% and 8% mass percentage (w/%w) PVA solutions were prepared by heating distilled water and adding a percentage weight of PVA to the water. The mixture was constantly stirred until the PVA particles were well hydrated upon which the holding vessel of PVA and water mixture was placed in an oven at 93°–95° Celsius for 7 hours. Small quantities of distilled water were added back into the solutions according to the amounts lost during preparation of the PVA solution. PVA solutions with different lower mass percentages can be subsequently produced by heating the master solutions and mixing in additional water.

2.2 PVA-C Preparation

PVA-C was prepared by completely freezing and thawing PVA solutions, which resulted in semi-translucent flexible gel. The solutions were placed in a room temperature (25° Celsius) chest freezer and cooled to -25° to -20° Celsius. After 12 hours of freezing at the aforementioned temperatures, the freezer was shut-off and its interior was allowed to rise back to room temperature over another 12 hour period. These freeze thaw cycles (FTC) were repeated as needed to vary the consistency of the produced PVA-C.

2.3 Brain Mold

The mold for our brain phantom was based on the polygonal surface mesh segmented from the left hemisphere of the Colin27 data set [1]. This mesh was then subtracted from a rectangular prism mesh to create a "negative" of the cortical surface, which was then used for stereolithographic printing.

We utilized the services of RedEye On Demand (Eden Prairie, MN, USA) for the fabrication of our brain phantom mold using the TangoPlus Polyjet Resin (FC-930) as the material. This clear rubber-like photopolymer is disposited layer by layer in order to produce a finished three dimensional object (See Fig. 1a). We found that the mold made using this material was able to accurately model the sulci and insular region of the cerebral hemisphere. It also has enough flexibility to allow demolding without damaging the PVA-C phantom and reverts itself to its original shape when deformed.

The bottom of a plastic tub was cut out and glued around the opening of the flexible rubber mold. This allows us to cast a base for our demolded phantom and limit its relative movement when placed and imaged in another plastic tub of the same size.

The mold component for the left ventricle of the phantom was constructed separately using silicone bathroom caulk. Layers of caulk approximately 2mm thick were applied to vellum traces from life-size printouts of segmented 2mm sagittal sections of the left ventricle. These layers were then assembled medially to laterally and aligned using cross-hairs on the printout traces to maintain placement accuracy of the sagittal sections and then covered with additional silicone rubber caulk to smooth the mold component surface.

2.4 Approximating Live Brain Texture

An array of PVA-C samples was prepared with either 1, 2, or 3 FTCs and 4%, 5%, 6%, or 8% PVA solutions, producing 12 different PVA-C samples. The samples were palpated at room-temperature by a gloved neurosurgeon who is familiar with the texture of human brain and brain pathologies.

The experienced neurosurgeon was then asked to rate each sample from 0 to 10 with the former being nothing like live brain and the latter being exactly like live brain, while noting whether the sample was softer or firmer than healthy brain tissue. The neurosurgeon was then asked to choose 2 of the samples which felt most like low grade gliomas.

The surgeon rated the PVA-C sample created from 6% PVA solution at 1 FTC as being similar to palpating the surface of a live brain and the 4% PVA at 3 FTC as being similar to palpating a low grade glioma.

We measured the elastic modulus of the 6% PVA 1 FTC PVA-C sample using a 3369 Dual Column Testing System (Instron, Norwood, MA, USA) and found that it has a value 4.6kPa, which is within the range found for human brain tissue [15].

2.5 Implants

To make our phantom useful for tests in image guidance and registration, structures created from various formulations of PVA-C were included into the phantom. Spherical registration marker were created using a harder PVA-C made from 8% PVA solution that has undergone 2 freeze-thaw cycles (FTC). These were molded using the containers for reflective passive spheres used in optical tracking and are approximately 12mm in diameter. A PVA-C "tumour" was also created using 4% PVA solution with 2 FTC and then implanted into the phantom. The tumour was molded using the plastic case from inside a Kinder surprise (Ferrero, Pino Torinese, Italy). When the phantom is completed, the markers and the tumour will each have undergone 3 FTC.

These PVA-C structures are skewered and suspended using 0.45mm monofilament fishing lines inside the phantom at the desired location. The left ventricle mold component was clamped and also suspended with fishing lines in a similar fashion.

Finally, we placed the inflatable head of a urinary catheter into the frontal lobe of the phantom and another in the medial portion of the phantom in the cast base of the mold. Each urinary catheter can be inflated with up to 10ml of water using a syringe in the manner described by Reinertsen and Collins [4] to vary the extent of deformation on the phantom.

3 Triple Modality Imaging Contrast

Commonly available chemicals were used to change the contrast of our phantom for imaging in US, MR, and CT. A PVA-C with the PVA concentration and FTC resembling textures similar to a living human cerebrum was chosen to be the base solution for dissolving the contrast enhancing chemicals.

To increase back-scattering of sound waves in US imaging, solutions containing talcum powder at 4%, 2%, 1%, and 0.5% weight of the base solution were mixed. Each of the samples was immersed in water and imaged with a Sonosite 180 Plus (Sonosite, Bothell, WA, USA) diagnostic ultrasound system tracked using a Stealth neurosurgical station (Medtronic, Minneapolis, MN, USA) and visually examined for contrast with the surrounding water and implanted PVA markers spheres.

For increasing phantom contrast in CT imaging, a powdered barium sulphate ($BaSO_4$) preparation used for colon enema (Guerbet Micropaque Colon, Guerbet, Villepinte, Île-de-France, France) was mixed into our PVA solutions. Solutions were prepared with 8%, 6%, 3%, and 1% weight $BaSO_4$ of the initial base solution. PVA-C samples were made from each of these solutions and imaged using a LightSpeed 16 VCT scanner (GE Healthcare, Little Chalfont, Buckinghamshire, UK).

To enhance the signal in T1 weighted images, copper sulphate ($CuSO_4$) was added to the PVA mixture in small quantities. Minute quantities of $CuSO_4$ dramatically increase the contrast of the PVA sample in T1 and T2 weighted images. To find an optimal concentration of $CuSO_4$, we prepared 0.2%, 0.1%,

0.05%, and 0.025% anhydrous $CuSO_4$ PVA-C samples and imaged them with T1 and T2 imaging sequences on a Verio 3T MR scanner (Siemens Healthcare, Erlangen, Germany).

We found that for the phantom brain tissue, PVA-C made with 6% PVA solution at 1FTC containing concentration of 2% $BaSO_4$, 0.025% $CuSO_4$, and 1% talcum as contrast agents worked well for CT, MR, and US, respectively. For our triple modality image markers, we found that a gel made from a 8% PVA solution with 5% $BaSO_4$, 0.2% $CuSO_4$, and 5% talcum as contrast agents worked well for CT, MR, and US, respectively. Although, $CuSO_4$ diffuses rather quickly out of the PVA-C, it did not affect the quality of the images greatly if they were acquired within the week when the phantom was built. The texture of the PVA-C did not change dramatically with the addition of these quantities of contrast agents.

4 Image Acquisition

Multimodality images with deformations were acquired for the phantom using the same models of MR, CT, and US imaging devices that we used for determining imaging contrast (See Fig. 1b,c). The phantom was first scanned in the MR with T1 weighted spin-echo (TR=668ms, TE=8.9ms, Flip Angle=70°, 1×1×3mm) and gradient-echo imaging (TR=1900ms, TE=3ms, Flip Angle=9°, 1mm isotropic), T2 weighted imaging (TR=6530ms, TE=840ms, Flip Angle= 150°, 1×1×3mm), proton density (PD: TR=6530ms, TE=9.4ms, Flip Angle=150°, 1×1×3mm), fluid attenuated inversion recovery (FLAIR: TR=5000ms, TE=273ms, TI=1800ms, Flip Angle=120°, 1mm isotropic), and 30 direction diffusion weighted (DWI: TR=9300ms, TE=94ms, Flip Angle=90°, 1×1×2mm) MR sequences. Fractional anisotropy, apparent diffusion coefficient, and trace weighted images were computed from the diffusion weighted images. A CT scan (491 axial slices at 1.25mm thickness) was then acquired for the phantom followed by the acquisition of a series of tracked B-mode US images (44 images of 4-6 sweeps at each 5.2cm and 7.1cm depth).

All of the images from each modality were acquired one after the other to ensure that the phantom is in the same state in the set of images. After each series of multimodality imaging, the phantom was deformed by inflating each of the two implanted urinary catheters in the phantom with 0ml, 5ml or 10ml of water through 5 rounds of inflations (See Fig. 2).

As well, images for super-resolution image processing were acquired by scanning the phantom with MP-RAGE T1 weighted gradient echo sequence 6 times at an isotropic resolution of 0.5mm with a slight displacement of the phantom less than 1cm between each acquisition. The phantom was then scanned using a T1 weighted spin echo sequence at 1mm isotropic resolution 7 times with its container filled with water and 7 times without the water, again with a slight displacement of the phantom less than 1cm between each acquisition.

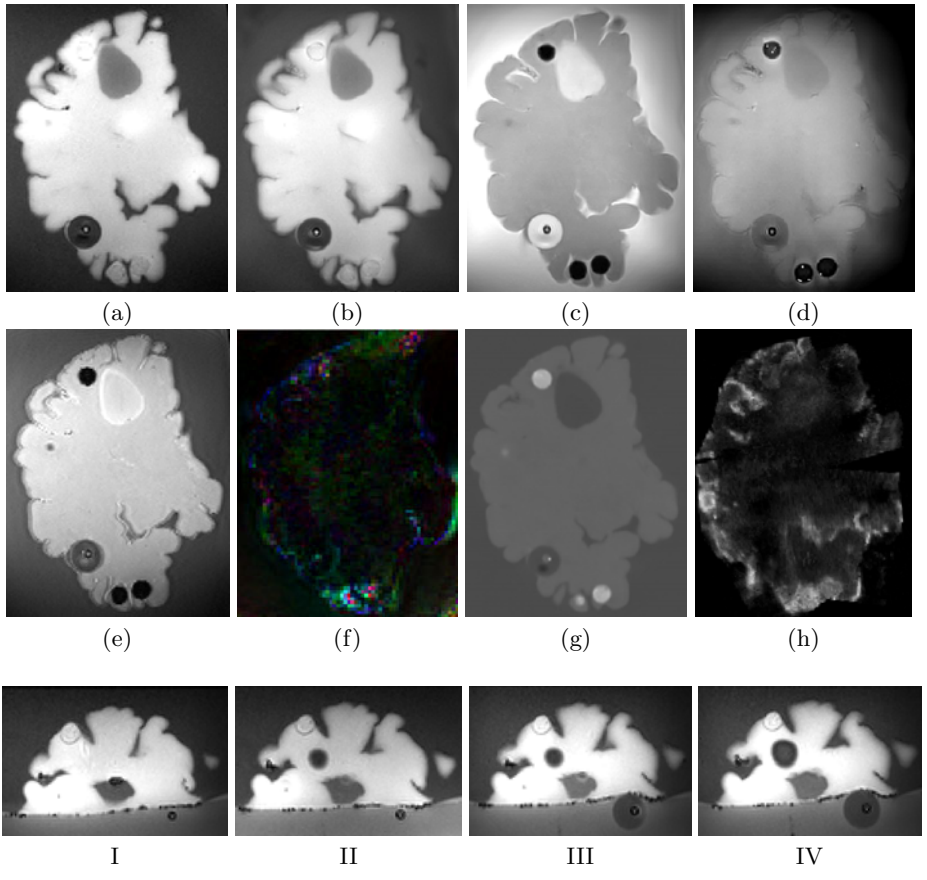


Fig. 2. A selection of PVA-C brain phantom images acquired using (a) MR T1-weighted gradient-echo (b) MR T1-weighted spin-echo (c) MR T2-weighted (d) MR PD (e) MR FLAIR (f) MR DTI colour map (g) CT (h) and the reconstructed US sweeps. Images I–IV shows the phantom imaged with T1-weighted gradient-echo at different inflations of the catheters.

5 Conclusion

An anatomically and mechanically realistic PVA-C brain phantom was created and imaged using MR, CT, and US. The images acquired from this phantom were then made publically available to the larger image processing community. We believe that the acquired multi-modal images be used for validation of many image processing techniques such as segmentation, super-resolution, image reconstruction, linear or non-linear registration, and denoising algorithms, using images acquired from one modality to act as the ground truth of another. The images are made available at: `http://pvabrain.inria.fr`

Aside from image processing, the formulation of our phantom material to approximate live cerebral brain tissue can be invaluable for improving implantation for deep-brain stimulators and simulating biopsy needle insertions. The accurate anatomy and texture of the brain phantom as well as the low cost of the starting materials can also make it useful as a tool in training medical professionals.

In order to further improve the phantom, we are currently in the process of developing better multimodal imaging spherical markers for higher contrast imaging, and finding better MR contrast agents that do not diffuse. We would also like to further characterize the multi-modal PVA-C developed in the work such that the physical and imaging property of the material can be better understood. Nevertheless, the presented methods represent an important step in the development of multimodality imageable tissue-like materials, along with techniques for creating anatomically accurate brain phantoms.

Acknowledgements. We would like to thank the Neurimage platform team for its financial support and their assistance in processing and acquiring the MR data, as well as Professor Olivier Le Guillou and Florence Le Dret at the chemistry department of the Institut National des Sciences Appliquées (INSA) de Rennes for allowing us to use their equipment for preparing solutions. We are also grateful to the Fonds québécois de la recherche sur la nature et les technologies (FQRNT) and the Collège Doctoral International (CDI) de l'Université européenne de Bretagne (UEB) Bourse de Mobilité Internationale Entrante for funding this project.

References

1. Holmes, C., Hoge, R., Collins, L., Woods, R., Toga, A., Evans, A.: Enhancement of MR images using registration for signal averaging. Journal of Computer Assisted Tomography 22(2), 324 (1998)
2. Surry, K.J.M., Austin, H.J.B., Fenster, A., Peters, T.M.: Poly(vinyl alcohol) cryogel phantoms for use in ultrasound and mr imaging. Physics in Medicine and Biology 49(24), 5529–5546 (2004)
3. Reinertsen, I., Descoteaux, M., Drouin, S., Siddiqi, K., Collins, D.L.: Vessel driven correction of brain shift. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 208–216. Springer, Heidelberg (2004)
4. Reinertsen, I., Collins, D.L.: A realistic phantom for brain-shift simulations. Medical Physics 33(9), 3234–3240 (2006)
5. Peppas, N.A.: Turbidimetric studies of aqueous poly(vinyl alcohol) solutions. Die Makromolekulare Chemie 176(11), 3433–3440 (1975)
6. Stauffer, S.R., Peppas, N.A.: Poly (vinyl alcohol) hydrogels prepared by freezing-thawing cyclic processing. Polymer 33(18), 3932–3936 (1992)
7. Peppas, N.A., Stauffer, S.R.: Reinforced uncrosslinked poly (vinyl alcohol) gels produced by cyclic freezing-thawing processes: a short review. Journal of Controlled Release 16(3), 305–310 (1991)
8. Chu, K.C., Rutt, B.K.: Polyvinyl alcohol cryogel: An ideal phantom material for MR studies of arterial flow and elasticity. Magnetic Resonance in Medicine 37(2), 314–319 (1997)

9. Duboeuf, F., Bbasarab, A., Liebgott, H., Brusseau, E., Delechartre, P., Vray, D.: Investigation of pva cryogel young's modulus stability with time, controlled by a simple reliable technique. Medical Physics 36(2), 656–661 (2009)
10. Mehrabian, H., Samani, A.: Constrained hyperelastic parameters reconstruction of pva (polyvinyl alcohol) phantom undergoing large deformation. In: Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling, vol. 7261 (March 2009)
11. Manohar, S., Kharine, A., van Hespen, J.C.G., Steenbergen, W., van Leeuwen, T.G.: Photoacoustic mammography laboratory prototype: imaging of breast tissue phantoms. Journal of Biomedical Optics 9(6), 1172–1181 (2004)
12. Fromageau, J., Gennisson, J.L., Schmitt, C., Maurice, R., Mongrain, R., Cloutier, G.: Estimation of polyvinyl alcohol cryogel mechanical properties with four ultrasound elastography methods and comparison with gold standard testings. IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control 54(3), 498–509 (2007)
13. Khaled, W., Neumann, T., Ermert, H., Reichling, S., Arnold, A., Bruhns, O.: Evaluation of material parameters of pva phantoms for reconstructive ultrasound elastography. In: IEEE Ultrasonics Symposium, pp. 1329–1332 (2007)
14. Jia, C., Kim, K., Kolias, T., Weitzel, W., Rubin, J., O'Donnell, M.: Left ventricular phantom with pulsatile circulation for ultrasound strain rate imaging. In: IEEE Ultrasonics Symposium, pp. 1317–1320 (2006)
15. Fallenstein, G., Hulce, V., Melvin, J.: Dynamic mechanical properties of human brain tissue. Journal of Biomechanics 2(3), 217–226 (1969)

Statistical Analysis of Structural Brain Connectivity

Renske de Boer[1,2], Michiel Schaap[1], Fedde van der Lijn[1], Henri A. Vrooman[1], Marius de Groot[1], Meike W. Vernooij[2,3], M. Arfan Ikram[2,3], Evert F.S. van Velsen[1,2], Aad van der Lugt[3], Monique M.B. Breteler[2], and Wiro J. Niessen[1,4]

[1] Biomedical Imaging Group Rotterdam, Departments of Radiology & Medical Informatics, Erasmus MC, Rotterdam, The Netherlands
[2] Department of Epidemiology, Erasmus MC, Rotterdam, The Netherlands
[3] Department of Radiology, Erasmus MC, Rotterdam, The Netherlands
[4] Imaging Science & Technology, Faculty of Applied Sciences, Delft University of Technology, Delft, The Netherlands

Abstract. We present a framework for statistical analysis in large cohorts of structural brain connectivity, derived from diffusion weighted MRI. A brain network is defined between subcortical gray matter structures and a cortical parcellation obtained with FreeSurfer. Connectivity is established through minimum cost paths with an anisotropic local cost function and is quantified per connection. The connectivity network potentially encodes important information about brain structure, and can be analyzed using multivariate regression methods. The proposed framework can be used to study the relation between connectivity and e.g. brain function or neurodegenerative disease. As a proof of principle, we perform principal component regression in order to predict age and gender, based on the connectivity networks of 979 middle-aged and elderly subjects, in a 10-fold cross-validation. The results are compared to predictions based on fractional anisotropy and mean diffusivity averaged over the white matter and over the corpus callosum. Additionally, the predictions are performed based on the best predicting connection in the network. Principal component regression outperformed all other prediction models, demonstrating the age and gender information encoded in the connectivity network.

1 Introduction

Both functional and anatomical connectivity of the brain are areas of increasing research interest. Functional connectivity is mainly established through functional magnetic resonance imaging (fMRI) while diffusion MRI has been applied for assessing anatomical or structural connectivity. Both types have been analyzed by modeling connectivity as a complex network and applying graph theory approaches to study the network topology. A review of graph theoretical analysis of complex brain networks is given in [1].

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 101–108, 2010.

In structural connectivity the network nodes represent brain regions and the connections are usually established through probabilistic or streamline tractography. Streamline tractography might be incapable of finding a connection in regions of uncertain directionality, due to e.g. crossing fibers or noise. To overcome this problem, probabilistic tractography was proposed in which multiple flow vectors are chosen from a distribution around the principal eigenvector of the diffusion tensor, e.g. [2,3]. Alternatively, a connection between two regions can be established through directional dependent minimum cost path methods [4,5,6,7]. These methods compute the minimum cost to get from a start region to another point in the image. The costs depend on both location and direction and can be defined based on the diffusion MRI data. Even though minimum cost path methods are closely related to probabilistic tractography, they contain no random factor and will therefore give reproducible results. This is an advantage when creating a connectivity network. Furthermore, minimum cost path methods find the globally optimal paths, except for some inaccuracy due to discretization. Probabilistic tractography is more likely to end at a local optimum because of limited flow vector sampling in a modeled distribution.

We present a new framework for statistical analysis of structural brain connectivity based on minimum cost paths. The connectivity network is established from diffusion weighted images (DWI) using the method previously proposed by Melonakos et al. [7]. The network nodes are based on segmentations and cortical parcellations obtained using FreeSurfer [8,9]. By quantifying the connectivity between the nodes, we construct a mean connectivity brain network that can be analyzed using multivariate statistics. The results can be used to study connectivity changes in e.g. aging, cognitive decline, and neurological or psychiatric disorders. As a proof of principle, we perform principal component regression in order to predict age and gender in a large dataset of aging subjects. We compare the results to predictions based on fractional anisotropy (FA) and mean diffusivity (MD) averaged over the white matter and the corpus callosum, and prediction based on the best predicting connection.

2 Framework

2.1 Connectivity

Minimum cost paths. The connectivity between two brain regions is defined by minimum cost paths. Start region R is connected with point $\mathbf{p}$, through the path Γ with minimum cumulative traveling cost. This cumulative cost $u(\mathbf{p})$ is given by

$$u(\mathbf{p}) = \min_{\Gamma} \int_0^L \psi(\mathbf{x}(s), \mathbf{x}'(s))ds \tag{1}$$

where s is the arc length along Γ; L is the length of Γ; $\mathbf{x}(s)$ is the position on the path; $\mathbf{x}'(s) = \frac{\nabla u}{||\nabla u||}$ is the unit local direction of the path; and $\psi(\mathbf{x}, \mathbf{v})$ is the local anisotropic cost function, defining the cost at position $\mathbf{x}$ in direction $\mathbf{v}$.

Local Cost Function. In order to have the minimum cost paths run through white matter bundles, a local cost function which is low on, and in the direction of, white matter tracts should be defined. To this end, different cost functions have been proposed. Some are based on the diffusion tensor model [4,6]. These methods are not suitable for modeling regions of multiple fiber populations e.g. in the event of crossing fibers. We choose a local cost function based on the set of acquired diffusion weighted images [5,7] and decide to use the local cost function proposed in [7].

$$\psi(\mathbf{x}, \mathbf{v}) = \left(\frac{S(\mathbf{x}, \mathbf{v})}{\int_{\mathbf{w} \perp \mathbf{v}} \frac{S(\mathbf{x}, \mathbf{w})}{S(\mathbf{x}, \mathbf{0})} d\mathbf{w}} \right)^3 \tag{2}$$

Where $S(\mathbf{x}, \mathbf{v})$ is the (interpolated) DWI value at position $\mathbf{x}$ and (interpolated) direction $\mathbf{v}$. $S(\mathbf{x}, \mathbf{0})$ is the value of the image without diffusion weighting (B0) at position $\mathbf{x}$. $S(\mathbf{x}, \mathbf{v})$ is low if the diffusion at position $\mathbf{x}$ in direction $\mathbf{v}$ is high, because of diffusion-related signal loss. Therefore, the costs are low if the diffusion is high in direction $\mathbf{v}$ compared to perpendicular directions.

Quantifying Connectivity. For every connection a value needs to be obtained depicting the connectivity between the connected nodes. It is possible to use different measures for quantifying connectivity. Equation 1 can be generalized to integrate any local measure, $f(\mathbf{x})$, from the start region R to point $\mathbf{p}$ over the minimum cost path, defined by $\psi(\mathbf{x}, \mathbf{v})$. Dividing the cumulative measure by the path length L yields a mean measure, $\bar{f}(\mathbf{x})$, over the minimum cost path.

$$\bar{f}(\mathbf{p}) = \frac{1}{L} \int_0^L f(\mathbf{x}(s)) ds \tag{3}$$

In this way it is possible to calculate e.g. the mean FA or mean MD over the minimum cost path. Although these measures are based on a tensor model, which has disadvantages as discussed before, the local cost function is not, and depends on both local anisotropy and diffusion. We use the mean cost, obtained by dividing cumulative cost by path length, as connectivity measure. Dividing by the length of the minimum cost path is necessary in order to correct for differences in head size and/or brain atrophy.

2.2 Construction of the Connectivity Network

To enable statistical analysis of brain connectivity maps, corresponding subcortical and cortical regions should be defined in all subjects. Hereto, the publicly available FreeSurfer software is used, which is capable of segmenting subcortical structures [8] and parcellating the cortex [9]. T1-weighted (T1w) scans are given as input for the FreeSurfer reconstruction pipeline. The resulting segmentation and cortical parcellation are transformed according to rigid registration of the T1w scan to the B0 diffusion image performed by Elastix [10]. Minimum cost paths are calculated as proposed by Melonakos et al. [7] and only performed within gray and white matter as defined by the FreeSurfer segmentation.

From the FreeSurfer segmentation, we use 17 subcortical structures succeedingly as start regions: the brain stem, and the left and right segmentations of thalamus, caudate nucleus, putamen, pallidum, hippocampus, amygdala, accumbens area and ventral diencephalon. As target regions we use the FreeSurfer cortical parcellation based on the Destrieux 2009 atlas, which divides the cortex in 75 regions per hemisphere, augmented by the 16 subcortial regions not currently used as start region. The connectivity between start and target region is represented by the minimum of the mean cost among all voxels in the target region. The resulting connectivity network consists of $2 \times 75 + 17 = 167$ nodes and $n = (167 - 1) \times 17 = 2822$ connections, which for m subjects combines into a $m \times n$ connectivity matrix.

2.3 Statistical Analysis

The connectivity matrix can be used to investigate changes in connectivity with e.g. normal aging, in cognitive decline, and in psychiatric disorders. As a proof of principle, we perform principal component regression (PCR) in an attempt to predict age and gender of our test subjects. For PCR we perform principal component analysis on the connectivity matrix. The first 15 principal components are used as input for multivariate linear regression to predict age and for multivariate logistic regression to predict gender. We compare our results to predictions based on five different univariate regression models. Two models are based on diffusion measures in the entire white matter, namely mean FA and MD (WM-FA and WM-MD). Additionally, two predictions are based on the regional measures of mean FA and mean MD in the corpus callosum (CC-FA and CC-MD). White matter and corpus callosum are defined by the FreeSurfer segmentation. For the fifth model, we perform a regression for every connection in the connectivity network. The connection with the smallest root mean squared deviation based on the training set, is used for the final prediction.

All experiments are performed in a 10-fold cross validation. For every fold, the regression coefficients are estimated on the training set and the prediction is evaluated on the test set. The alternative prediction models use the same subdivision of subjects for the cross validation. Age prediction is evaluated by mean absolute difference ($|\Delta|$) between predicted and actual age. For the prediction of gender, the percentage of correctly predicted subjects is reported.

In population studies of the elderly, gender may not be distributed evenly over all ages. In that case, if both age and gender relate to the prediction model variable, it is necessary to correct for one when predicting the other. This correction is performed by linear regression with the confounding variable as input, and the model variable as output. The residuals are used for prediction.

3 Results

Imaging data from the Rotterdam Scan Study [11], acquired in 2005-2006, were used for the evaluation of the method. Scans were obtained on a 1.5 T GE scanner using an 8-channel head coil. The DWI scanning protocol had a b-value

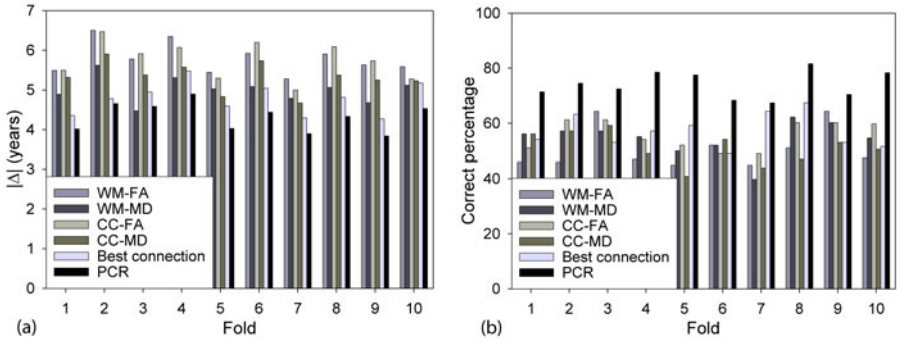


Fig. 1. Mean absolute difference in age prediction (a) and percentage correctly predicted gender (b) for all models at all folds

of 1000 s/mm^2 in 25 non-collinear directions, and one volume was acquired without diffusion weighting. Voxel sizes were $0.8 \times 0.8 \times 3.5$ mm. Head motion and Eddy current corrections of the DWI were performed with FDT, part of FSL [12]. FDT was also used to fit the tensor for calculation of the FA and MD images. The 3D T1w images had voxels sizes of $0.49 \times 0.49 \times 0.8$ mm.

Subjects with cortical infarcts, artifacts in any of their scans or FreeSurfer errors were excluded from the analysis. The remaining 979 subjects had a mean age of 68.5 $\pm$ 7.4 years (range 59.0 - 96.7) and consisted of 469 men and 510 women.

Figure 1a shows the age prediction error for the different regression models per fold. Lower $|\Delta|$ indicates better results. The PCR model resulted in the lowest $|\Delta|$ for nine out of ten folds, while the FA-based models showed overall the highest $|\Delta|$. Table 1 shows the results averaged over all folds. PCR on the connectivity matrix improved age prediction with at least 0.7 years compared to the global and regional measures. It also showed an improvement of 0.5 years compared to the best predicting single connection. The 15 principal components explained on average 65.8% of the variance in the gender corrected data.

Figure 1b shows the gender prediction accuracy for all models and folds. PCR outperformed the other models on all folds. Table 1 shows that, while the prediction based on the global and regional diffusion measures was close to random, PCR on the connectivity matrix obtained a gender prediction accuracy of 74.1%. The 15 principal components explained on average 61.6% of the variance in the age corrected data.

Figure 2 shows the connections running between the right putamen start region and all target regions in one subject. These connections are obtained by tracing back in the direction of the lowest cumulative cost, starting at the target voxel. By combining the 15 principal components with the PCR coefficients, it is possible to obtain the per connection regression coefficients. The connections in Fig. 2 are colored according to their regression coefficient for gender prediction as calculated based on the training set of the tenth fold.

Table 1. Per model, mean absolute difference in years in predicted age and percentage correctly predicted gender averaged over all folds

	Age prediction $\|\Delta\|$ (years)	*Gender prediction* Correct (%)
WM-FA	5.8	50.8
WM-MD	5.0	54.4
CC-FA	5.8	55.8
CC-MD	5.3	51.1
Best connection	4.8	57.2
PCR	4.3	74.1

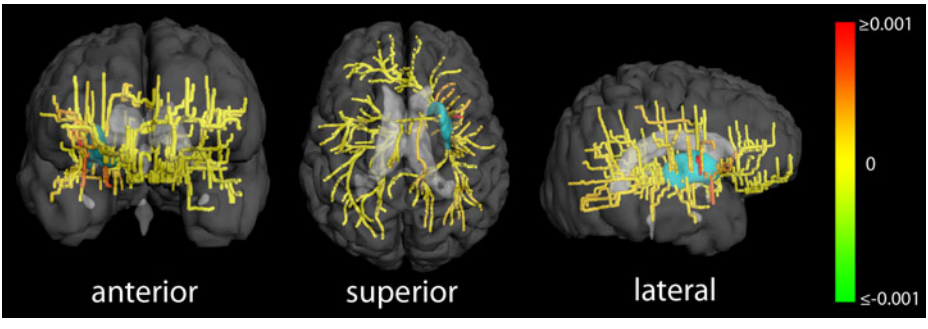


Fig. 2. Subject specific connections starting from the right putamen (in cyan). Connections are colored according to their regression coefficients for gender prediction.

4 Conclusion and Discussion

We present a new framework for structural connectivity analysis in large datasets of diffusion weighted brain MRI. Connectivity is established through minimum cost paths with an anisotropic local cost function based on DWI. Using brain regions defined by FreeSurfer, a connectivity network is obtained that can be analyzed using multivariate regression methods. As a proof of principle we performed PCR in order to predict age and gender. WM-FA and WM-MD are capable of predicting age with a mean absolute difference of 5.8 and 5.0 years respectively. PCR of the connectivity matrix reduces the prediction error to 4.3 years, suggesting that some of the network connections contain more information regarding age than others. The tight age distribution of the subjects makes prediction a difficult task. Predicting the mean age or, in case of a skewed distribution, the median age can already be quite accurate in tight distributions. The mean absolute difference for prediction of the mean age of the training set is 6.0 years; the median age reduces this difference to 5.6 years. Both results are close to the results of the global and regional diffusion measure models, but PCR

shows a clear improvement. On gender prediction, PCR improves the correctly predicted percentage from close to random for the global and regional diffusion measures to 74.1%.

PCR predicts more accurate than the best predicting connection for both age and gender. It could be argued that this improvement is caused by the use of 15 versus one regression variable, or by the aggregation of connections into groups. The latter could result in averaging of noise due to redundancy in the data. To test the additional advantage of using principal components over using the mean of groups of connections, we randomly assigned all connections into 15 groups, which are used as input for multivariate linear regression. This model results in a mean absolute difference in predicted age of 4.9 years and a percentage of correctly predicted gender of 62.7%. These results are comparable or better than the best predicting connection. However, the results are not better than the PCR results, which shows the benefits of the use of principal components.

Minimum cost path methods will always find a connection between two regions. This is an advantage when constructing a connectivity matrix in which corresponding connections need to be found in all subjects. As a result, the proposed framework differs from existing approaches in that the analysis is carried out on all connections instead of a subset. Furthermore, there is no need for parameters that determine which connections are retained in the analysis [1]. The disadvantage, however, is that it might suggest biologically implausible direct connections between regions. It is important to keep in mind that the connections that are found may also represent indirect connections, connecting two regions through a third.

Finally, the presented whole-brain analysis method is especially suited to study differences in connectivity in populations, where previously proposed tractography-based analyses are mainly directed at studying the topology of the structural network. Robinson et al. also perform statistical analysis of brain connectivity, but they use probabilistic tractography based on a tensor model and classify their subjects in two wide range age groups [13]. In [14] they perform their tractography based on a two-component partial volume model and classify in two strongly contrasting age groups. We do not fit any model to the DWI data and predict age as a continuous variable. The prediction of age or gender of a person is of course not very relevant in research and clinical practice. The performed experiment can, however, be used to assess which connections have the largest contribution to the prediction. These connections contain the most information regarding age or gender. In future work, we will add the cortical parcellation regions as start regions, creating a connectivity network that includes cortico-cortical connections.

In conclusion, we present a framework for construction of a structural brain connectivity network that potentially can be used to study brain changes in e.g. aging, neurodegenerative disease or psychiatric disorders. As a proof of principle, we perform PCR in order to predict age and gender based on a network of connections between subcortical structures and cortical regions. PCR outperforms

the predictions based on global and regional averaged FA and MD and the best predicting single connection, demonstrating the value of the information encoded in the connectivity network.

References

1. Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nat. Rev. Neurosci. 10, 186–198 (2009)
2. Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., et al.: Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077–1088 (2003)
3. Parker, G.J.M., Haroon, H.A., Wheeler-Kingshott, C.A.M.: A framework for a streamline-based probabilistic index of connectivity (PICo) using a structural interpretation of MRI diffusion measurements. J. Magn. Reson. Imaging 18, 242–254 (2003)
4. Jackowski, M., Kao, C.Y., Qiu, M., Constable, R.T., Staib, L.H.: White matter tractography by anisotropic wavefront evolution and diffusion tensor imaging. Med. Image Anal. 9, 427–440 (2005)
5. Pichon, E., Westin, C.F., Tannenbaum, A.R.: A Hamilton-Jacobi-Bellman approach to high angular resolution diffusion tractography. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 180–187. Springer, Heidelberg (2005)
6. 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. Inf. Process Med. Imaging 20, 346–358 (2007)
7. Melonakos, J., Pichon, E., Angenent, S., Tannenbaum, A.: Finsler active contours. IEEE Trans. Pattern Anal. Mach. Intell. 30, 412–423 (2008)
8. Fischl, B., Salat, D.H., van der Kouwe, A.J.W., et al.: Sequence-independent segmentation of magnetic resonance images. NeuroImage 23(suppl. 1), S69–S84 (2004)
9. Fischl, B., van der Kouwe, A., Destrieux, C., et al.: Automatically parcellating the human cerebral cortex. Cereb. Cortex 14, 11–22 (2004)
10. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.: Elastix: A toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging 29, 196–205 (2010)
11. Hofman, A., Breteler, M.M.B., van Duijn, C.M., et al.: The Rotterdam Study: 2010 objectives and design update. Eur. J. Epidemiol. 24, 553–572 (2009)
12. Woolrich, M.W., Jbabdi, S., Patenaude, B., et al.: Bayesian analysis of neuroimaging data in FSL. NeuroImage 45, S173–S186 (2009)
13. Robinson, E.C., Valstar, M., Hammers, A., Ericsson, A., Edwards, A.D., Rueckert, D.: Multivariate statistical analysis of whole brain structural networks obtained using probabilistic tractography. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 486–493. Springer, Heidelberg (2008)
14. Robinson, E.C., Hammers, A., Ericsson, A., Edwards, A.D., Rueckert, D.: Identifying population differences in whole-brain structural networks: A machine learning approach. NeuroImage, 910–919 (2010)

Maximum A Posteriori Estimation of Isotropic High-Resolution Volumetric MRI from Orthogonal Thick-Slice Scans

Ali Gholipour, Judy A. Estroff, Mustafa Sahin,
Sanjay P. Prabhu, and Simon K. Warfield

Departments of Radiology and Neurology, Childrens Hospital Boston,
and Harvard Medical School, 300 Longwood Ave., Boston, MA 02115,
http://www.crl.med.harvard.edu/

Abstract. Thick-slice image acquisitions are sometimes inevitable in magnetic resonance imaging due to limitations posed by pulse sequence timing and signal-to-noise-ratio. The estimation of an isotropic high-resolution volume from thick-slice MRI scans is desired for improved image analysis and evaluation. In this article we formulate a maximum a posteriori (MAP) estimation algorithm for high-resolution volumetric MRI reconstruction. As compared to the previous techniques, this probabilistic formulation relies on a slice acquisition model and allows the incorporation of image priors. We focus on image priors based on image gradients and compare the developed MAP estimation approach to scattered data interpolation (SDI) and maximum likelihood reconstruction. The results indicate that the developed MAP estimation approach outperforms the SDI techniques and appropriate image priors may improve the volume estimation when the acquired thick-slice scans do not sufficiently sample the imaged volume. We also report applications in pediatric and fetal imaging.

Keywords: MAP estimation, super-resolution, volume reconstruction.

1 Introduction

Thick slice image acquisitions are sometimes inevitable in magnetic resonance imaging (MRI) due to pulse sequence timing requirements and the need to maintain high signal-to-noise-ratio (SNR). Such scans are typically performed for T2-weighted and diffusion weighted imaging in a variety of applications including brain, lung, and heart imaging, and fetal and neonatal MRI. Thick-slice scans are acquired in single shot fast spin echo (SSFSE) imaging of fetuses, neonates, and pediatric patients who may move in the scanner [1]. SSFSE slices are acquired in a fraction of a second, thus freezing the motion of the subject. Nevertheless, due to thick slice acquisitions necessary to maintain SNR and the inter-slice motion artifacts, these images do not appropriately reflect the 3D anatomy.

The reconstruction of a high-resolution volumetric image from thick slice scans is desired for enhanced image analysis and improved evaluation. Inter-slice reconstruction has been previously addressed in [2] based on an iterative back

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 109–116, 2010.

projection reconstruction algorithm, where multiple shifted thick-slice scans provide dense sampling of the imaged object. In more recent studies [1], [3], [4] high-resolution volumes have been reconstructed from fast slice scans of moving subjects. These scans are affected by inter-slice motion thus iterations of slice-to-volume registration and scattered data interpolation (SDI) have been used in these studies for iterative motion estimation and volume reconstruction.

Nevertheless, scattered data interpolation techniques do not provide a mathematical framework to justify that the estimated high-resolution volume is a minimum error representation of the imaged object given the acquired scans. In this article we have formulated a general mathematical framework based on a maximum a posteriori (MAP) estimation algorithm for high-resolution volume reconstruction. Inspired by the recent advances in super-resolution image reconstruction [5], [6], the developed MAP estimation approach relies on a slice acquisition model and minimizes a cost function of the error norm between the estimated volume and the acquired slices.

In addition, the MAP estimation approach allows the incorporation of prior image models for volume reconstruction, which is critical when the number of thick-slice scans is limited and the slice thickness is significantly larger than the matrix resolution. Under certain conditions the developed MAP estimation approach simplifies to regularized MLE reconstruction, thus the main contribution in this article is the development and performance analysis of MAP volume estimation and comparison to SDI and non-regularized MLE solutions.

As such, we focus on simple image priors based on image gradients for performance analysis of high-resolution volume reconstruction using the formulated MAP estimation approach. We limit our evaluation to the reconstruction of isotropic brain volumes from a limited number of orthogonal thick-slice scans, but the results can be generalized to similar applications. Our evaluation involves quantitative analysis using synthetic digital brain phantom images, and applications in pediatric and fetal MRI. We compare the MAP estimation technique with the B-Spline SDI approach in [1] and a non-regularized MLE solution, and evaluate the effect of image priors under different scanning conditions.

2 Methods

In order to formulate the volume estimation problem in a super-resolution framework we need to establish a slice acquisition model, which describes how the acquired slices are obtained from the imaged object. The following slice acquisition model is considered in this study:

$$\mathbf{y}_k = \mathbf{D}_k\mathbf{B}_k\mathbf{S}_k\mathbf{M}_k\mathbf{x} + \mathbf{v}_k; k = 1, ..., n \tag{1}$$

where $\mathbf{y}_k$ is the vector of the voxels of the k^{th} 2D slice with slice thickness Δs_k and uniform in-plane spacing of $\Delta\rho_k$; $\mathbf{x}$ is a vector of the desired reconstructed image voxels in the lexicographical order with isotropic spacing of $\Delta\rho$; $\mathbf{v}_k$ is the residual noise vector, n is the number of slices obtained from N scans, $\mathbf{M}_k$ is the matrix of motion parameters, $\mathbf{S}_k$ is a matrix representing the slice selection

profile, $\mathbf{B}_k$ is a blur matrix representing the point spread function (PSF) of the MRI signal acquisition process, and $\mathbf{D}_k$ is a down-sampling matrix.

On the basis of Equation (1) the imaged object goes through geometric and signal operations, including motion, slice selection and signal averaging, PSF blur, and resampling, to generate the acquired slices. Assuming that all the matrix operations, including the motion parameter matrices are known in Equation (1), this equation can be written in a simple linear form like $\mathbf{y}_k = \mathbf{W}_k\mathbf{x} + \mathbf{v}_k$, where $\mathbf{W}_k = \mathbf{D}_k\mathbf{B}_k\mathbf{S}_k\mathbf{M}_k$. These linear equations can be augmented to form a large linear matrix equation like $\mathbf{y} = \mathbf{W}\mathbf{x} + \mathbf{v}$. Super-resolution volume reconstruction is the inverse problem of finding $\mathbf{x}$ given the acquired slices $\mathbf{y}_k$. The classical solution to this linear inverse problem can be obtained through maximum likelihood estimation (MLE).

The MAP estimation is considered as a generalization of MLE and is written based on the conditional probability density function (PDF) of the acquired slices $\mathbf{y}_k$ given the estimated volume $\hat{\mathbf{x}}$ as well as the prior information about the PDF of the estimated volume, i.e. $\Pr(\hat{\mathbf{x}})$:

$$\mathbf{x}_{\text{MAP}} = \arg\max_{\mathbf{x}} \left[\log\Pr(\mathbf{y}_k|\hat{\mathbf{x}}) + \log\Pr(\hat{\mathbf{x}})\right] \tag{2}$$

The MAP solution depends on the probability functions. Here we assume that the noise residuals (error samples) are drawn from Gaussian distributions with mean of zero and standard deviation of σ_k. Therefore:

$$\Pr(\mathbf{y}_k|\hat{\mathbf{x}}) = \prod_i \frac{1}{\sigma_k\sqrt{2\pi}} \exp\left(-\frac{(\hat{y}_k(i) - y_k(i))^2}{2\sigma_k^2}\right) \tag{3}$$

where $y_k(i)$ are the samples from the acquired slices $\mathbf{y}_k$, and $\hat{y}_k(i)$ are the samples from the estimated slices $\hat{\mathbf{y}}_k = \mathbf{W}_k\hat{\mathbf{x}} + \mathbf{v}_k$. The error samples are defined by $e_k(i) = \hat{y}_k(i) - y_k(i)$, and the error vector is defined as $\mathbf{e}_k = \mathbf{W}_k\hat{\mathbf{x}} - \mathbf{y}_k$.

Various image priors may be used. The simplest form involves an exponential function that is quadratic in the voxel values of $\mathbf{x}$, i.e. $\Pr(\hat{\mathbf{x}}) = \exp(-\hat{\mathbf{x}}^{\mathrm{T}}\mathbf{Q}\hat{\mathbf{x}})$; where $\mathbf{Q}$ is a symmetric, positive definite matrix. Here we use $\mathbf{Q} = \mathbf{C}^{\mathrm{T}}\mathbf{C}$ where $\mathbf{C}$ is the gradient magnitude image operation. Assuming independent slice acquisitions the log-likelihood of the conditional PDF in Equation (3) is the sum of the l_2-norm of the error vectors over all the slices. Consequently the maximization of the log likelihood function results in the following minimization problem:

$$\hat{\mathbf{x}}_{MAP} = \arg\min_{\mathbf{x}} \sum_{k=1}^{n} \|\mathbf{W}_k\hat{\mathbf{x}} - \mathbf{y}_k\|_2^2 + \lambda\|\mathbf{C}\hat{\mathbf{x}}\|_2^2 \tag{4}$$

The augmented matrix $\mathbf{W}$ in the linear inverse problem is very large and the classical solution through pseudo-inverse is prohibitive. Instead we use a steepest descent iterative minimization approach. The iterative solution of Equation (4) based on image operators shown in Equation (1) is written as:

$$\hat{\mathbf{x}}^{n+1} = \hat{\mathbf{x}}^n + \alpha \sum_{k=1}^{n} \mathbf{M}_k^{\mathbf{T}}\mathbf{S}_k^{\mathbf{T}}\mathbf{B}_k^{\mathbf{T}}\mathbf{D}_k^{\mathbf{T}}(\mathbf{y}_k - \mathbf{D}_k\mathbf{B}_k\mathbf{S}_k\mathbf{M}_k\hat{\mathbf{x}}^n) - \lambda\mathbf{C}^{\mathrm{T}}\mathbf{C}\hat{\mathbf{x}}^n, \tag{5}$$

where α is the step size in the direction of the gradient. The matrices $\mathbf{D}_k$, $\mathbf{B}_k$, $\mathbf{S}_k$, and $\mathbf{M}_k$ and their transposes are exactly interpreted as corresponding image operators. $\mathbf{D}_k$ is defined as a resampling operation. $\mathbf{B}_k$ is defined as the convolution with a Gaussian kernel resembling the point spread function (PSF) of the MRI signal acquisition process. $\mathbf{C}$ is implemented as a gradient magnitude image operation. λ is a weighting coefficient.

The slice selection profile $\mathbf{S}_k$ is defined based on the slice selection process. For an arbitrary slice select direction defined by the normal vector of the slice plane equation, the following geometrical equation is obtained for the voxels of slice k (defined by a vector $\boldsymbol{r}$) in the slice selection process: $|\boldsymbol{\mu}_{sk}.\boldsymbol{r} - s_{0k}| < \Delta s_k/2$; where Δs_k is the slice thickness, and s_{0k} is the distance of the slice from the origin. $\boldsymbol{\mu}_{sk}$ specifies the slice (or slice-selection) orientation and is interpreted as the normal vector of the slice plane equation. The normal vector can be obtained in the physical coordinate system based on the so-called direction cosines rotation matrix. Consequently the $\mathbf{S}_k$ operation is implemented as rigid 3D rotation with the rotation matrix directly obtained from direction cosines matrix, and its transpose is the inverse (transpose) of the direction cosines matrix.

The motion matrix $\mathbf{M}_k$ is implemented as a 6-DOF 3D rigid transformation (including three rotations and three translations). Motion estimation and super-resolution volume reconstruction are considered as separable problems. Therefore in the presence of inter-slice motion, iterations of motion correction and volume reconstruction are performed to find $\mathbf{M}_k$ and $\mathbf{x}$, respectively. Motion correction can be performed through slice-to-volume registration [1], [3], or based on slice intersections [4]. This is not a subject of interest in this article; in order to focus on the performance analysis of volume reconstruction we assume that there is no motion or the motion is known (accurately corrected) in our experiments.

3 Results

3.1 Quantitative Evaluation

Quantitative evaluation and comparison of the algorithms was carried out using digital brain phantom (DBP) images obtained from the *Brainweb* database [7]. Thick slice scans in the axial, coronal and sagittal slice select directions were synthetically generated from the high-resolution DBP images by applying operations based on Equation (1). The in-plane resolution of the synthetic scans was 1 mm and various slice thicknesses were examined between 2 to 8 mm.

Since a reference high-resolution volume is available for the validation dataset (i.e. the original DBP images), the accuracy of reconstruction can be measured quantitatively. Two measures are used here: Mean Absolute Error (MAE) and Peak Signal to Noise Ratio (PSNR). MAE is defined as the mean absolute differences of the voxel intensity values between the reference volume and the reconstructed volume. PSNR is defined in the logarithmic decibel (dB) scale as

$20\log_{10}(\mathrm{MAX}/\sqrt{\mathrm{MSE}})$, where MAX is the maximum possible voxel intensity value (4096 in our experiments) and MSE is the mean square error of the voxel intensity values between the reference volume and the reconstructed volume. Lower MAE and higher PSNR indicate more accurate reconstruction.

Fig. 1 shows the MAE and PSNR values computed as a function of the slice thickness of synthesized thick-slice input scans for different volume reconstruction techniques. Four techniques have been considered: AVE is the simplest one and is based on averaging the input scans resampled to the space of the desired high-resolution volumetric image. SDI is a 3-level BSpline SDI approach based on [1], MLE is a non-regularized MLE obtained from the MAP formulation by setting $\lambda = 0$, and MAP is the MAP estimation with $\lambda = 0.01$. This value was chosen experimentally as a normalization factor between the reconstruction error image and the gradient magnitude of the estimated volume.

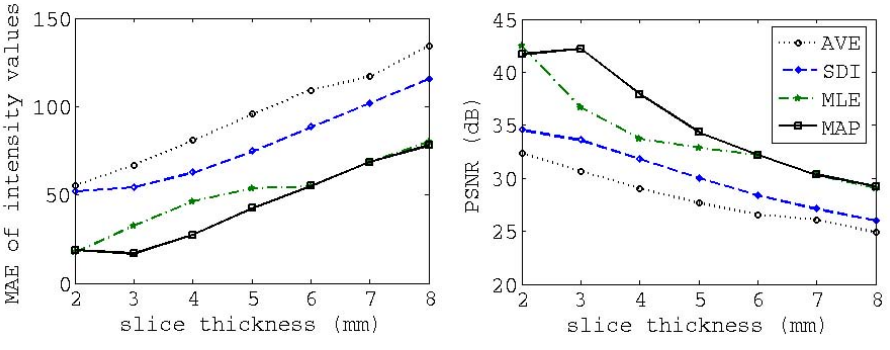


Fig. 1. MAE and PSNR between the ground truth DBP volume and the reconstructed volumes as a function of slice thickness. The measures have been compared for 4 techniques: AVE for averaging the resampled input scans, SDI for BSpline SDI, MLE for non-regularized MLE, and MAP for the developed MAP estimation approach.

The results in Fig. 1 indicate that all the techniques perform better than simple averaging. The developed MAP estimation approach outperforms SDI, and is generally more accurate than MLE. The comparison of MAP and MLE at slice thickness 2 mm indicates that if the slice thickness is not much larger than the in-plane resolution, the prior image model may not provide improvements. On the other hand, when the high-resolution volume space is not densely sampled by the thick-slice scans (i.e. due to large slice thickness and limited number of orthogonal scans), image priors significantly improve the reconstruction accuracy. This is observed for the slice thickness values between 3 to 5 mm. Finally, when the slice thickness is too large (i.e. more than six times larger than the in-plane resolution) image priors in the form of image gradients may not help too much. Improvement of volume estimation is difficult in these cases due to fundamental performance limits in super-resolution reconstruction [8].

3.2 Application to Pediatric and Fetal MRI

The first application is the estimation of isotropic high-resolution volumes from thick-slice T2-weighted TSE scans of pediatric patients who underwent clinical brain MRI for the evaluation of tuberous sclerosis. As part of the imaging procedure, two T2-weighted TSE volumes (one axial and one coronal) were acquired for each patient using a Siemens Trio 3-Tesla scanner. TSE imaging was performed with $TR = 14070$ ms, $TE = 89$ ms, matrix size of 512×512, slice thickness of 2 mm, and in-plane resolutions between 0.4 to 0.5 mm. The scanning protocol also involved a high-resolution T1-weighted (T1W) MPRAGE acquisition with $TR = 2530$ ms, $TE = 3.39$ ms, and isotropic resolution of 1 mm.

The acquired TSE scans, as well as the reconstructed volumes and the T1W MPRAGE volume of a 3-year-old child are shown in Fig. 2. The reconstructed volumes in this case have a high isotropic resolution of 0.5 mm^3, which is four times better than the slice thickness and two times better than the MPRAGE volume. Visual inspection indicates that the MAP estimated volume is much sharper and has a better contrast as compared to the SDI estimated volume.

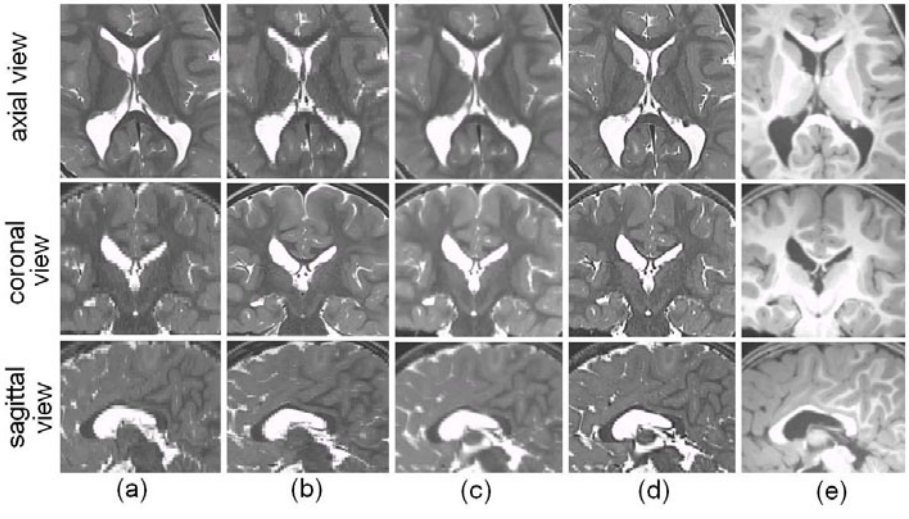


Fig. 2. Application of the volume reconstruction algorithms to T2-weighted TSE images of a 3-year old child: Two thick-slice TSE scans were acquired and used for reconstruction in the (a) axial and (b) coronal directions; (c) and (d) are the volumes reconstructed using the SDI approach and the MAP estimation approach, respectively, and (e) is the acquired high-resolution T1W MPRAGE volume

We use two sets of measures for comparing the accuracy of volume reconstructions. First we compute the similarity of the reconstructed T2W volumes to the acquired high-resolution T1W volume. Normalized mutual information (NMI) is appropriate in this case as it quantifies the nonlinear relationship between

the intensity values of images with different contrast sources. We also use two sharpness (focus) measures: M1 (the intensity variance measure) and M2 (the energy of image gradient measure). Both measures are monotonic and robust to noise [9]. The variance measure is calculated as the sum of square differences (SSD) between each voxel intensity value and the mean image intensity value. M2 is computed by integrating the magnitude of image gradient at all voxels.

The NMI measure computed as the similarity of T2W volume to the acquired T1W volume was 1.61, 1.71, and 2.00 for the AVE, SDI, and MAP estimation techniques respectively. This indicates that from an information-theoretic viewpoint the intensity values of the MAP estimated volume better match with those of the reference T1W volume. The computed M1 & M2 sharpness measures were 39092 & 6.7×10^{11}, 41671 & 1.0×10^{12}, and 43139 & 1.1×10^{12} for the AVE, SDI, and MAP estimation techniques respectively. This indicates that the sharpest volumes were obtained from the MAP estimation approach.

The second application of this technique is for fetal MRI. Iterative inter-slice motion correction and volume reconstruction was performed here. Clinical fetal MRI scans were obtained using a 1.5-T TwinSpeed Signa system and an 8-channel phased-array cardiac coil for pregnant patients with diagnosed or suspected cases of fetal anomalies after diagnostic ultrasonography. The input scans involve multiple SSFSE acquisitions in the fetal sagittal, axial and coronal planes with slice thickness between 3 to 5 mm, and in-plane resolution of 0.7 to 0.8 mm.

Fig. 3 shows an example of volumetric fetal brain MRI reconstruction. Note that the reconstructed volume clearly reflects the underlying continuity of tis-

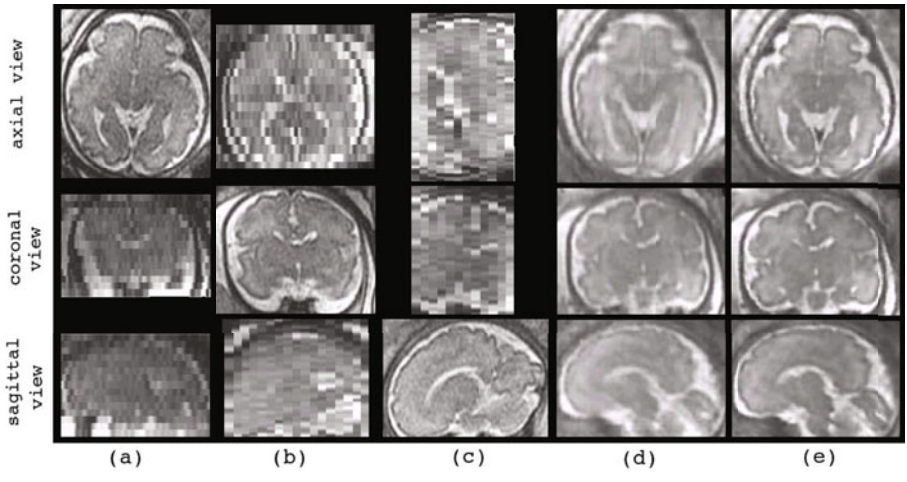


Fig. 3. Application of the volume reconstruction algorithms to a 31.43 week fetus: Three of the six acquired SSFSE scans in fetal axial, coronal, and sagittal planes are shown in (a) to (c) respectively. (d) and (e) show the volumes reconstructed with isotropic resolution of 0.8 mm^3 using the SDI approach and the MAP estimation approach, respectively. Note that coherent tissue boundaries present in all three planes of the MAP estimated volume but not in the out-of-plane views of the original scans

sue structural boundaries in all three planes, whereas the original acquisitions exhibit discontinuous tissue boundaries in the out-of-plane views due to the effect of partial volume averaging. We examined 15 fetal brain MRI datasets and computed sharpness measures. The average improvement in the M1 and M2 sharpness measures with respect to the AVE reconstructed volumes were 8% and 20% for the SDI, and 12% and 42% for the MAP estimated volumes, respectively.

4 Conclusion

We have developed a MAP estimation approach for the reconstruction of isotropic high-resolution volumetric MRI from thick-slice orthogonal scans. This formulation is based on a slice acquisition model, minimizes a cost function of an error norm between the acquired thick-slice scans and the reconstructed volume, and provides a framework for the incorporation of image priors. The results indicate that the MAP estimation approach outperforms the scattered data interpolation techniques, and image priors result in improved accuracy when the slice thickness is 3 to 5 times larger than the in-plane resolution. In addition to fetal and pediatric MRI, this approach can be used in many other MRI applications.

Acknowledgments. This research was supported in part by NIH grants R01 RR021885, R01 GM074068, R01 EB008015, and R43 MH086984.

References

1. Jiang, S., Xue, H., Glover, A., Rutherford, M., Rueckert, D., Hajnal, J.: MRI of moving subjects using multislice snapshot images with volume reconstruction (SVR): application to fetal, neonatal, and adult brain studies. IEEE Transactions on Medical Imaging 26(7), 967–980 (2007)
2. Greenspan, H., Oz, G., Kiryati, N., Peled, S.: MRI inter-slice reconstruction using super-resolution. Magnetic Resonance Imaging 20(5), 437–446 (2002)
3. Rousseau, F., Glenn, O., Iordanova, B., Rodriguez-Carranza, C., Vigneron, D., Barkovich, J., Studholme, C.: Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images. Acad. Radiol. 13, 1072–1081 (2006)
4. Kim, K., Habas, P., Rousseau, F., Glenn, O., Barkovich, A., Studholme, C.: Intersection based motion correction of multi-slice MRI for 3d in utero fetal brain image formation. IEEE Transactions on Medical Imaging 29(1), 146–158 (2010)
5. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction: a technical overview. IEEE Signal Processing Magazine 20(3), 21–36 (2003)
6. Farsiu, S., Robinson, M., Elad, M., Milanfar, P.: Fast and robust multiframe super resolution. IEEE Transactions on Image Processing 13(10), 1327–1344 (2004)
7. Collins, D., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., Evans, A.: Design and construction of a realistic digital brain phantom. IEEE Transactions on Medical Imaging 17(3), 463–468 (1998)
8. Robinson, D., Milanfar, P.: Statistical performance analysis of super-resolution. IEEE Transactions on Image Processing 15(6), 1413–1428 (2006)
9. Subbarao, M., Choi, T., Nikzad, A.: Focusing techniques. Journal of Optical Engineering 32, 2824–2836 (1992)

Change Detection in Diffusion MRI Using Multivariate Statistical Testing on Tensors

Antoine Grigis[1,2], Vincent Noblet[1], Félix Renard[1,2], Fabrice Heitz[1], Jean-Paul Armspach[2], and Lucien Rumbach[2]

[1] University of Strasbourg, CNRS, UMR 7005, LSIIT, France
[2] University of Strasbourg, CNRS, FRE 3280, LINC-IPB, France
grigis@unistra.fr

Abstract. This paper presents a longitudinal change detection framework for detecting relevant modifications in diffusion MRI, with application to Multiple Sclerosis (MS). The proposed method is based on multivariate statistical testings which were initially introduced for tensor population comparison. We use these methods in the context of longitudinal change detection by considering several strategies to build sets of tensors characterizing the variability of each voxel. These testing tools have been considered either for the comparison of tensor eigenvalues or eigenvectors, thus enabling to differentiate orientation and diffusivity changes. Results on simulated MS lesion evolutions and on real data are presented. Interestingly, experiments on an MS patient highlight the ability of the proposed approach to detect changes in non evolving lesions (according to conventional MRI) and around lesions (in the normal appearing white matter), which might open promising perspectives for the follow-up of the MS pathology.

1 Introduction

The automated detection of relevant changes in longitudinal Magnetic Resonance Imaging (MRI) sequences is crucial for medical diagnosis, follow-up and prognosis. The core problem is to identify image regions that are significantly different between two successive images. Because of its widespread interest and the large number of applications in diverse domains, change detection has been the topic of much interest. A review of the literature can be found in [1]. In conventional MRI, various methods have already been proposed to detect changes [2], [3], [4]. But few works have addressed change detection in Diffusion Tensor Imaging (DTI). Some previous works have addressed change detection in scalar images, characterizing diffusion properties, such as the Fractional Anisotropy or the Mean Diffusivity. These works rely either on statistical parametric testing, using the Generalized Likelihood Ratio Test (GLRT) [5], or on non parametric testing, for instance permutation testing [6]. A comparison of standard statistical testing on Diffusion Weighted (DW-) images is presented in [7]. In [8], the GLRT has been extented to diffusion tensor images, but the approach does not take into account the positive definite nature of matrices. This constraint has been

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 117–124, 2010.

considered in [9] where tensor test statistics have been developed for population comparison. We propose to use this method for longitudinal change detection. To this end, we consider stategies to build sets of tensors characterizing the *variabily* of each voxel. These strategies are based on the variability existing in the DW-images, or in the spatial neighborhood of the considered voxel, or a combination of these two. Based on the tensor model, we derive testing tools for the comparison of tensor eigenvalues and eigenvectors, thus enabling to differentiate orientation and diffusivity changes. The proposed framework is based on the following processing pipeline: 1- Preprocessing of the two DTI acquisitions, 2- Generation of the two tensor populations to be compared and 3- Multivariate statistical testing between the two sets of tensors.

2 Proposed Framework

2.1 Preprocessing

Eddy current distortion correction. Each set of DW-images may be misaligned due to eddy current distortions. Thus, a correction is achieved by affinely registering each slice of the DWI data set onto the corresponding slice of the T2 image (without diffusion weighting).

Registration. Change detection methods generally require the two images to be accurately registered. Thus, an affine transformation is estimated between the two DWI sets, considering the mutual information between the T2 images. Since an affine transformation may not always be sufficient to compensate for all the undesired global differences that may be present between the two acquisitions, the registration is then refined with the deformable method described in [10].

Interpolation and reorientation. An interpolation method is required to resample the warped tensor image. It has been shown in [11] that interpolation methods for tensor image can lead to significantly different results according to the chosen metric. In particular, the Log-Euclidean metric seems to be prefered to the Euclidean metric because of the swelling effect induced by the latter. Therfore, we use a third order B-Spline interpolation method in the Log-Euclidean space. Then, the Preservation of Principal Direction (PPD) reorientation strategy described in [12] is applied to preserve the orientation and shape information of the warped tensor image.

2.2 Generation of the Tensor Sets Reflecting the Local Variability

In this section, we investigate several strategies to generate a set of tensors characterizing the variability of each voxel, based on the variability existing in the DW-images, or in the spatial neighborhood of the considered voxel, or a combination of these two. Tensors are estimated using the standard least squares approach. Since this procedure does not guarantee the positive definiteness of the tensors (which is a required property for using Log-Euclidean metrics), negative eigenvalues are set to an arbitrary small positive value.

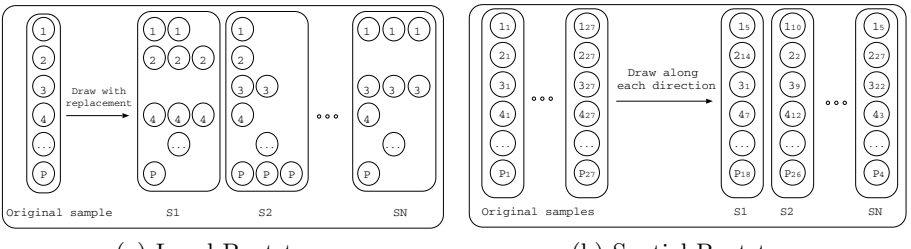
Local Bootstrap. DTI acquisitions are composed of P gradient directions, P being usually greater than six, thus leading to an overdetermined linear system to estimate the tensor. Using a bootstrap strategy enables us to learn the distribution reflecting tensor estimation uncertainty based on the DW-images [13]. To this end, we generate at each voxel N bootstrap samples by randomly selecting P DW-signal with replacement, and preserving the gradient directions [13] (Fig.1a). The random draw with replacement of the gradient directions amounts actually to associate different weights to each gradient direction for tensor estimation. Using the local bootstrap on the two registered data sets enables to generate two populations of tensors.

Spatial Neighborhood. An alternative idea is to learn the tensor distribution at each voxel by considering all the tensors in a surrounding user-defined spatial neighborhood. By this way, we make the implicit assumption that tensors follow a constant model on this neighborhood and that the observed variability is the consequence of random effects. Learning tensor distribution at each voxel by considering all the tensors in a surrounding spatial neighborhood is based on the commonly made assumption of a constant piecewise model [1], [2]. The limitations of this model are well known, in particular here, at the interface between different tissues, but the model is known to be robust and efficient. This method has the advantage to be computationally cheap. However, it does not take into account the information carried out by the DW-images.

Spatial Bootstrap. The spatial bootstrap is a combination of the local bootstrap and the spatial neighborhood methods. From a surrounding box with a user-defined size, we generate at each voxel N bootstrap samples by drawing for each gradient direction one DW-signal of the neighborhood (Fig.1b).

2.3 Multivariate Statistical Testing on Tensors

Log-Euclidean metric. Using Euclidean metrics, such as the Frobenius distance, may raise some problem when considering symmetric positive definite



(a) Local Bootstrap (b) Spatial Bootstrap

Fig. 1. Schematic representation of the local and spatial ($3 \times 3 \times 3$ neighborhood) bootstrap for P gradient directions. 1_{27} represents the DW-signal of the 27th voxel of the neighborhood according to the first gradient direction.

(SPD) matrices since it does not define a vector space on SPD matrices. A simple and computationally efficient way to build a vector space on SPD matrices is to consider the Log-Euclidean metric [11]. The Log-Euclidean metric corresponds to the Euclidean metric on the logarithm of matrices. By this way, the matrices with null or negative eigenvalues are at an infinite distance to any SPD matrices. The logarithm of an SPD matrix D is obtained as $L = Log(D) = U Log(\Lambda) U^T$, where Λ and U are the matrices derived from the standard spectral decomposition, containing respectively the eigenvalues and the eigenvectors of D. Then, the Log-Euclidean distance between two SPD matrices D_1 and D_2 can be defined as the Euclidean distance between their logarithms:

$$d^2(D_1, D_2) = \|Log(D_1) - Log(D_2)\|^2 \tag{1}$$

According to this metric, an estimator of the mean $\bar{D}$ of a set of N tensors D_i is given by the exponential of the arithmetic log-tensors mean, i.e. $\bar{D} = exp(\bar{L})$, with:

$$\bar{L} = argmin_{\Sigma} \sum_{i=1}^{N} \|Log(D_i) - \Sigma\|^2 = \frac{1}{N} \sum_{i=1}^{N} Log(D_i) \tag{2}$$

Statistical tests. Many statistical tests rely on the normal distribution. Considering the multivariate normal distribution for the SPD matrices has the drawback to associate matrices with negative or null eigenvalues with a non null probability. To circumvent this limitation, Schwartzman [9] suggests to model the matrix logarithms with the multivariate normal distribution, which comes to model the SPD matrices with a Log-normal distribution. Based on this model, it is possible to derive statistical tests on tensors eigenvalues and eigenvectors. We consider two populations of N_1 and N_2 tensors respectively. Under the assumption that the tensor logarithms of the two populations follow the normal distributions $L_1 \sim \mathcal{N}(M_1, \sigma^2 Id)$ and $L_2 \sim \mathcal{N}(M_2, \sigma^2 Id)$, the Maximum Likelihood Estimates of M_1, M_2, and σ^2 are respectively $\bar{L_1}$, and $\bar{L_2}$ computed according to Eq. 2, and:

$$\hat{\sigma}^2 = \frac{1}{6(N_1 + N_2 - 2)} \left[\sum_{i=1}^{N_1} tr\left(L_{1i} - \bar{L_1}\right)^2 + \sum_{i=1}^{N_2} tr\left(L_{2i} - \bar{L_2}\right)^2 \right] \tag{3}$$

We consider a test [9] that evaluates whether the two populations of diffusion tensors have similar eigenvalues, but possibly different eigenvectors. Let D_1, U_1 and D_2, U_2 be the matrices derived from standard spectral decomposition, and containing respectively the eigenvalues and eigenvectors of M_1 and M_2. The test, based on the log-likelihood ratio under hypotheses $H0 : D_1 = D_2$ vs $H1 : D_1 \neq D_2$ is:

$$T = \frac{(N_1 + N_2 - 2) N_1 N_2}{3(N_1 + N_2)^2 \hat{\sigma}^2} tr\left[(\Lambda_1 - \Lambda_2)^2\right] \tag{4}$$

with Λ_1 and Λ_2 the eigenvalue matrices of $\bar{L_1}$ and $\bar{L_2}$, respectively.

Another case of interest is to test whether the two population of diffusion tensors have similar eigenvectors, i.e the same orientation. When we combine the two populations, we derive the mean tensor $\bar{L}$ of the new population from Eq. 2. We compute then the eigen-decomposition of this latter, and note Λ the eigenvalue matrix. The proposed test, based on the log-likelihood ratio under hypotheses $H0 : U_1 = U_2$ vs $H1 : U_1 \neq U_2$, treating $D_1 = D_2 = D$ as a nuisance parameter is:

$$T = \frac{2(N_1 + N_2) - 3}{6\hat{\sigma}^2} tr\left[\left(\frac{N_1\Lambda_1 - N_2\Lambda_2}{N_1 + N_2}\right)^2 - \Lambda^2\right] \quad (5)$$

3 Results

Experiments on synthetic lesions. We considered two successive DTI acquisitions of the same subject acquired on a 3.0T MRI scanner with 30 encoding gradients (b-value of 1000 s/mm^2). By this way, the differences between the two scans are only due to the acquisition noise and distortion. The image dimensions are $128 \times 128 \times 40$ and the spatial resolution is $1.8 \times 1.8 \times 3.5mm^3$. A synthetic lesion is simulated in one of these scans as follows. We consider a lesion mask, located in the white matter, and delineated by a physician to ensure its location and shape to be clinically relevant. Inside the lesion mask, we modify uniformly either the diffusivity in the principal direction, i.e. the principal eigenvalue (application of a multiplicative factor $k \in]1, 2[$), or the tensor orientations (rotation of the ellipsoid canonical xyz system by $\alpha \in]0, \pi/8]$). The criterion used to compare the different methods is the area under Receiver Operating Characteristic (ROC) curves. In a ROC curve, the true positive rate (or sensitivity) is plotted in function of the false positive rate (or 1-specificity) for different cut-off points. A test that allows a perfect discrimination is characterized by a ROC plot that passes through the upper left corner (i.e. an area of one). The Table 1 summarizes the results for the three investigated strategies: S (Spatial), LB (Local Bootstrap), and SB (Spatial Bootstrap). For all the experiments, we consider a $3 \times 3 \times 3$ spatial neighborhood for the SB and S strategies, and $N = 100$ bootstrap samples for LB and SB strategies. For both diffusivity and orientation modifications, the two methods based on the spatial neighborhood stand out, thus pointing out the prominent part of the spatial information for reliable change detection.

Table 1. Areas under the ROC curve for eigenvalues and eigenvectors tests (MS simulation: see text)

eigenvalues:

k	S	LB	SB
1.2	0.713	0.202	0.724
1.4	0.902	0.417	0.927
1.6	0.960	0.743	0.982
1.8	0.976	0.787	0.991

eigenvectors:

α	S	LB	SB
$\pi/32$	0.814	0.590	0.726
$2 * \pi/32$	0.871	0.756	0.844
$3 * \pi/32$	0.870	0.822	0.850
$4 * \pi/32$	0.871	0.877	0.873

Experiments on an MS patient. We considered two brain DTI acquisitions of an MS patient acquired on a 3.0T MRI scanner with 33 encoding gradients (b-value of 1000 s/mm^2) at two different times ($t_2 - t_1 = 15\ months$). The images dimensions are $256 \times 256 \times 34$ and the spatial resolution is $1 \times 1 \times 3mm^3$. Examples of detection maps obtained with the three proposed approaches are presented in Fig.2. Each potential approach gives different results for the variance (Eq.3), and mean tensors estimation (Eq.2). For instance (Fig.2: var-LB), the DW-signal variability is greater in areas of high diffusivity (e.g. cerebrospinal fluid), and much lower in highly structured areas (e.g. white matter). On average, considering the spatial neighborhood gives larger values for the variance except in the cerebrospinal fluid (Fig.2: var-S). As a consequence, the bootstrap yields less false detections in the ventricles. The local bootstrap generates change detection maps sensitive to small variations since the variance values at each voxel are smaller (Fig.2: eig-LB). Directly using the spatial information after the tensors estimation, achieves satisfactory results for a low computational burden. A spatial bootstrap estimation benefits from the interesting properties of the local bootstrap (i.e., direct modeling of the DW-signal), and reduces false detections thanks to the spatial neighborhood information (Fig.2: SB).

A visual inspection of the results by an expert helps us to verify that anatomical changes are also observable in diffusion imaging. Then other changes are observed, that may provide information of a different kind. In Fig.2: eig, two

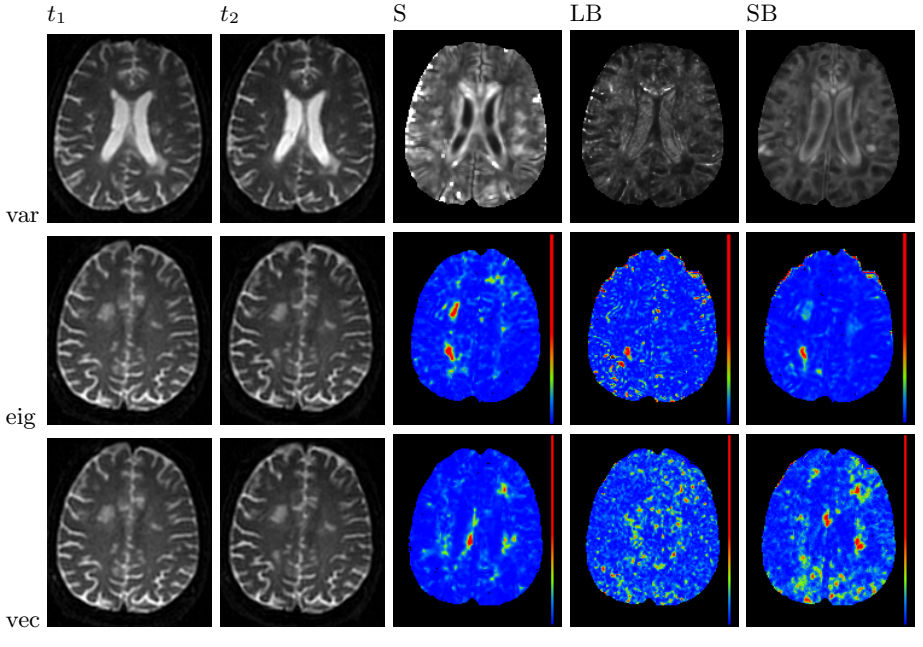


Fig. 2. Variance (var), and detection maps obtained for the eigenvalues (eig) and eigenvectors (vec) test statistics with the three different strategies for tensor population generation

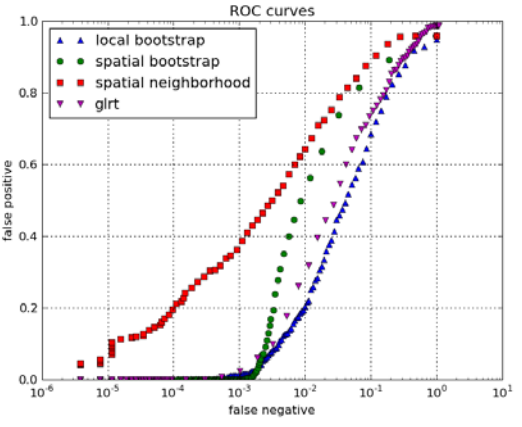


Fig. 3. ROC curves (semilogx plot): longitudinal change detection results for the three proposed generation of tensor sets and the GLRT

prominent lesions are detected. Both detections correspond to existing lesions (in anatomical images), but according to the expert, only the bottom one is changing. The top lesion is of widespread interest, since considering the anatomical images, it does not seem to evolve during the time elapsed between the two acquisitions. Then in Fig.2: vec, if it was obvious that for MS patients we have to detect variations in the eigenvalues map (demyelination Fig.2: eig), we also highlight the fact that the orientation is significantly changing. Notice also that the proposed framework detects white matter modification near the ventricles corresponding to physiological process known as leukoaraiosis.

ROC curves are used to quantitatively compare the three strategies, Fig.3. We compare the results with a reference detection map obtained by a manual segmentation of evolving lesions on the T2 images by an expert. Thus, ROC plots 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 anatomical images). The proposed framework is compared with the GLRT framework introduced in [8], which differs in the following points. During the preprocessing, we directly resample the DTI rather than going through the DW-signal. On top of that, the GLRT does not take into account the positive definite nature of matrices. From Fig.3, we highlight the interesting behavior of the proposed test statistic compared to the GLRT. The separability property (eigenvalues/eigenvectors) is of much interest.

4 Conclusion

Clinical disability for MS patients are localized in the white matter, and are driven by a significant modification of the perpendicular diffusivity. As expected, the eigenvalues test statitistic is relevant to detect such modifications. The com-

parison of three strategies for the generation of the two tensor sets to be compared has demonstrated the superiority of methods based on a spatial neighborhood. Nonetheless, the spatial information introduces a smoothing effect on the detection, while the bootstrap based approaches enable a better spatial localization. We also notice significant orientation modifications during longitudinal analysis. In the future we want to investigate if some correlation exists between MS disease evolution and these pathological modifications, and try to combine both information as a predictive tool for MS.

References

1. Radke, R.J., Andra, S., Al-Kofahi, O., Roysam, B.: Image change detection algorithms: a systematic survey. IEEE Trans. Image Process 14(3), 294–307 (2005)
2. Bosc, M., Heitz, F., Armspach, J.-P., Namer, I., Gounot, D., Rumbach, L.: Automatic change detection in multimodal serial MRI: application to multiple sclerosis lesion evolution. NeuroImage 20, 643–656 (2003)
3. Rey, D., Subsol, G., Delingette, H., Ayache, N.: Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis. In: Kuba, A., Sámal, M., Todd-Pokropek, A. (eds.) IPMI 1999. LNCS, vol. 1613, pp. 154–157. Springer, Heidelberg (1999)
4. Thirion, J.-P., Calmon, G.: Deformation analysis to detect and quantify active lesions in three-dimensional medical image sequences. IEEE Trans. Med. Imaging 18(5), 429–441 (1999)
5. Boisgontier, H., Noblet, V., Heitz, F., Rumbach, L., Armspach, J.-P.: An automatic method for change detection in serial DTI-derived scalar images. In: Workshop MIAMS - MICCAI (2008)
6. 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 (2008)
7. Boisgontier, H., Noblet, V., Heitz, F., Rumbach, L., Armspach, J.-P.: Statistical detection of longitudinal changes between apparent diffusion coefficient images. Application to multiple sclerosis. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 959–966. Springer, Heidelberg (2009)
8. Boisgontier, H., Noblet, V., Heitz, F., Rumbach, L., Armspach, J.-P.: Generalized likelihood ratio tests for change detection in diffusion tensor images. In: ISBI, pp. 811–814 (2009)
9. Schwartzman, A.: Random ellipsoids and false discovery rates: statistics for diffusion tensor imaging data. PhD thesis, Stanford University (June 2006)
10. Noblet, V., Heinrich, C., Heitz, F., Armspach, J.-P.: Retrospective evaluation of a topology preserving non-rigid registration method. Medical Image Analysis 10(3), 366–384 (2006)
11. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Geometric means in a novel vector space structure on symmetric positive-definite matrices. SIAM Journal on Matrix Analysis and Applications 29(1), 328–347 (2007)
12. Alexander, D.C., Pierpaoli, C., Basser, P.J., Gee, J.C.: Spatial transformations of diffusion tensor magnetic resonance images. IEEE Trans. Med. Imaging 20(11), 1131–1139 (2001)
13. Pajevic, S., Basser, P.J.: Parametric and non-parametric statistical analysis of DT-MRI data. Journal of Magnetic Resonance 161(1), 1–14 (2002)

Increasing Power to Predict Mild Cognitive Impairment Conversion to Alzheimer's Disease Using Hippocampal Atrophy Rate and Statistical Shape Models*

Kelvin K. Leung[1,2], Kai-Kai Shen[3,4], Josephine Barnes[2], Gerard R. Ridgway[1,2], Matthew J. Clarkson[1,2], Jurgen Fripp[3], Olivier Salvado[3], Fabrice Meriaudeau[4], Nick C. Fox[2], Pierrick Bourgeat[3], and Sébastien Ourselin[1,2]

[1] Centre for Medical Image Computing, University College London, WC1E 6BT, UK
[2] Dementia Research Centre, UCL Institute of Neurology, London, WC1N 3BG, UK
[3] Australian eHealth Research Centre, CSIRO ICT Centre, Australia
[4] Université de Bourgogne, Le2i UMR CNRS 5158, France

Abstract. Identifying mild cognitive impairment (MCI) subjects who will convert to clinical Alzheimer's disease (AD) is important for therapeutic decisions, patient counselling and clinical trials. Hippocampal volume and rate of atrophy predict clinical decline at the MCI stage and progression to AD. In this paper, we create p-maps from the differences in the shape of the hippocampus between 60 normal controls and 60 AD subjects using statistical shape models, and generate different regions of interest (ROI) by thresholding the p-maps at different significance levels. We demonstrate increased statistical power to classify 86 MCI converters and 128 MCI stable subjects using the hippocampal atrophy rates calculated by the boundary shift integral within these ROIs.

1 Introduction

The clinical onset of Alzheimer's disease is insidious and progressive, with the cognitive ability of a patient changing slowly from normal to severely impaired over many years. Many subjects are diagnosed with mild cognitive impairment (MCI) when they have measurable memory deficits but do not fulfill AD criteria. While around 30% of all patients diagnosed with MCI progress to AD, some remain stable and others revert to normal [1]. There is much interest in the early diagnosis of AD: identifying those MCI subjects who will progress to clinical AD (MCI converters) from those who remain stable (MCI stable). This information is valuable for making therapeutic decisions, counseling patients and planning

* This work was undertaken at UCL/UCLH which received a proportion of funding from the Department of Health's NIHR Biomedical Research Centres funding scheme. The Dementia Research Centre is an Alzheimer's Research Trust Co-ordinating centre. KKL and MC are supported by TSB grant M1638A, NCF is funded by the Medical Research Council (UK). JB is supported by an Alzheimer's Research Trust (ART, UK) Research Fellowship partly supported by the Kirby Laing Foundation.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 125–132, 2010.

clinical trials. Large scale studies such as the Alzheimer's Disease Neuroimaging Initiative (ADNI) (www.adni-info.org) were established to collect imaging and clinical data in order to assess and compare biomarkers of disease progression and early diagnosis of AD.

The hippocampus is one of the earliest structures affected in AD, and hippocampal atrophy on magnetic resonance imaging (MRI) has been shown to be a marker of AD pathology. Hippocampal atrophy is also predictive of clinical decline at an MCI stage and even presymptomatically in familial AD. Recent research has shown that hippocampal volume, shape variation and atrophy rate may differentiate MCI converters from stable subjects [2,3], and predict the conversion to AD [4]. In addition, regions (or subfields) in the hippocampus have been shown to have different atrophy patterns in AD, with inward deformations occurring in the CA1 and subiculum subfields in AD, while CA3,4 and dentate gyrus subfields remaining relatively intact [5]. Greater atrophy of the CA1 region is found in MCI converters [2]. Therefore, a region of interest (ROI) approach which identifies separately different sub-regions of the hippocampus may be more sensitive to study changes in hippocampal volumes.

A common method to identify ROIs is based on the statistical differences between different diagnostic groups. Hua et al. [6] recently showed that the sample size of a hypothetical AD clinical trial can be reduced by performing the analysis on ROIs based on voxels with significant atrophy rates over time ($p < 0.001$) between baseline and 1-year follow-up scans within the temporal lobes in a non-overlapping training set of 22 AD patients. Other researchers have used ROIs in the hippocampus identified using atrophy or surface maps based on statistical differences between MCI converters and stable subjects to demonstrate the ability to separate MCI converters and stable subjects [2,3]. In particular, Morra et al. [7] created an average surface map for each diagnostic group by geometrically averaging the surface maps of the subjects within that group. The local atrophy within the hippocampus is estimated by calculating the radial size of the hippocampus along its long axis. P-maps are then generated by comparing the local atrophy at each point in the hippocampus between different diagnostic groups.

1.1 Contributions

The main contribution of this paper is the proposal to identify ROIs in the hippocampus based on the statistical differences in its shape between normal controls (NC) and AD subjects using statistical shape models, and then quantify the atrophy rates within these local ROIs using the boundary shift integral (BSI). We demonstrate increased statistical power to classify MCI converters and stable subjects using atrophy rates within these ROIs.

2 Method

2.1 Image Data

We downloaded from the ADNI website (www.adni-info.org) pre-processed baseline and 12-month repeat volumetric T1-weighted MR scans acquired using

1.5T scanners of 334 subjects (60 NC, 214 MCI and 60 AD: mean(SD) age 76.5(4.8), 75.0(7.3) and 75.1(6.6)). MCI subjects were then subdivided into two subgroups (86 converters and 128 stable: mean(SD) age 74.8(7.4) and 75.2(7.2)) based on their follow-up clinical diagnoses determined up to 36 months after baseline. Representative imaging parameters were TR = 2400ms, TI = 1000ms, TE = 3.5ms, flip angle = 8°, field of view = 240 × 240mm and 160 sagittal 1.2mm-thick-slices and a 192 × 192 matrix, or 180 sagittal 1.2mm-thick-slices with a 256 × 256 matrix. The images are pre-processed by the standard ADNI image processing pipeline, which included post-acquisition correction of gradient warping, B1 non-uniformity correction, intensity non-uniformity correction and phantom based scaling correction.

Hippocampal segmentations at baseline and 12 months were downloaded from the ADNI website. The segmentations were calculated by using a non-linear warping technique from a template aided by the placement of manual landmarks [8]. The technique (referred to as SNT) is commercially available from Medtronic Surgical Navigation Technologies (Louisville, CO) and has been validated in elderly subjects including MCI and AD patients [9].

2.2 Method Overview

Statistical shape models (SSM) of the hippocampus using these baseline SNT hippocampal segmentations were built from a training set of 120 randomly selected subjects (60 NC and 60 AD). A Hotelling's T^2 test was performed on the position of each landmark to identify regions of significant shape difference between NC and AD. The resulting p-maps were then mapped onto the 214 MCI subjects. Thresholding of the p-maps at different significance levels was used to generate different ROIs in the hippocampus of the MCI subjects. We then tested whether the quantification of hippocampal atrophy rates using BSI within these ROIs increased the ability to classify MCI converters and stable subjects.

2.3 Statistical Shape Model

Two SSMs were built on the training set of 120 (NC+AD) subjects for left and right hippocampus respectively. The hippocampal surfaces generated by marching cube algorithm were first rigidly aligned using the iterative closest point algorithm. The correspondence between shape landmarks was established by a groupwise optimization of minimal description length with fluid regularization in the shape image [10]. The hippocampal shapes were then aligned by the Procrustes analysis via similarity transformations. A principal component analysis was then performed and the components were chosen to accounts for 98% of total variations in the population (assuming that the last 2% would be noise), so that the shapes in the training set can be expressed as

$$\mathbf{x}_i = \bar{\mathbf{x}} + \mathbf{W}\mathbf{b}_i \tag{1}$$

where $\mathbf{W}$ is the matrix consisting of eigenvectors of the covariance matrix, and the elements in vector $\mathbf{b}_i$ are the parameters for the i-th shape. For each

landmark on the hippocampal surface, a Hotelling's T^2 test was used in order to assess the differences between NC and AD.

The marching cube generated MCI hippocampal surfaces were smoothed using a windowed sinc function to remove the aliasing artifact due to voxelization. The SSM was deformed to fit the smoothed target surface $\mathcal{S}_{\text{MCI}}$, minimizing a distance metric between the SSM generated surface and the target. The distance between two surfaces $\mathcal{S}(\mathbf{x})$ and $\mathcal{S}(\mathbf{y})$ generated by point sets $\mathbf{x}$ and $\mathbf{y}$ can be defined in a symmetrical manner as

$$d_S(\mathcal{S}(\mathbf{x}), \mathcal{S}(\mathbf{y})) = \sum_{\mathbf{p}\in\mathbf{x}} d_p(\mathbf{p}, \mathcal{S}(\mathbf{y})) + \sum_{\mathbf{q}\in\mathbf{y}} d_p(\mathbf{q}, \mathcal{S}(\mathbf{x})), \tag{2}$$

where $d_p(\mathbf{p}, \mathcal{S})$ is the Euclidean distance from the point to the closest point on surface $\mathcal{S}$.

Thus we can fit the SSM to the target surface $\mathcal{S}_{\text{MCI}}$ by the optimization of parameters

$$(\mathcal{T}_{\mathbf{y}}, \mathbf{b}_{\mathbf{y}}) = \arg\min_{(\mathcal{T},\mathbf{b})} d_S(\mathcal{S}(\mathcal{T}(\bar{\mathbf{x}} + \mathbf{W}\mathbf{b})), \mathcal{S}_{\text{MCI}}) \tag{3}$$

using Powell's algorithm, where $\mathcal{T}$ is a similarity transformation with 7 degrees of freedom. For each landmark on the SSM-generated surface, we propagated its p-value to the closest point on the surface $\mathcal{S}_{\text{MCI}}$. Thus we can have a p-map for any given hippocampal surface.

Different ROIs on the surface of the hippocampus were obtained by thresholding the p-maps using $p < 0.05$, $p < 0.01$, $p < 0.005$, $p < 0.001$, $p < 0.0005$ and $p < 0.0001$.

2.4 Hippocampal Atrophy Using Boundary Shift Integral

We used the hippocampal boundary shift integral (HSBI) to quantify the atrophy rate between the baseline and repeat scans [11,12]. After aligning the hippocampus in the repeat images to the baseline image using a 6-degree-of-freedom registration, HBSI was calculated over the hippocampal boundary region (given by the exclusive OR region of the binary 1-voxel dilated and 1-voxel eroded hippocampal regions) using a double intensity window approach [13,14][1]. A double intensity window was included for the HBSI calculation in order to capture boundary shift at both the hippocampus–CSF border, and the hippocampus–WM border.

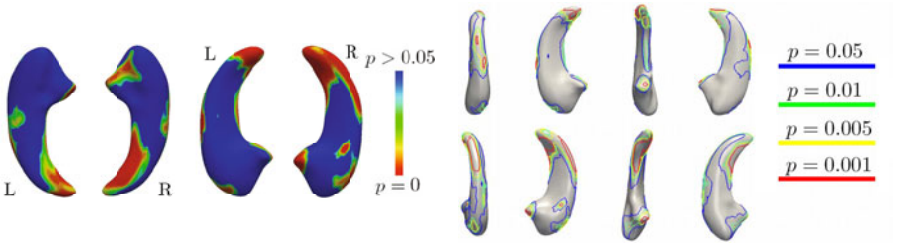
To quantify the hippocampal atrophy rates within the ROIs identified using SSM, we calculated HBSI over the AND region of the hippocampal boundary region and the ROIs dilated by 1 voxel. The hippocampal atrophy rates were then calculated by dividing HBSI by the baseline hippocampal volume and normalized by the scan interval.

Cohen's d and the area under the curve from receiver operating curves (ROC) were used to assess the relative statistical power of hippocampal atrophy rates calculated using different ROIs to classify MCI converters and stable subjects.

[1] BSI source code is availabe at http://sourceforge.net/projects/bsintegral/

3 Results

Fig. 1(a) shows the p-map of the differences between NC and AD. Regions with low p-value (in red) correspond to the medial and lateral aspects of the hippocampal head and tail. The iso-contours of p-values outlining regions with different significance levels are shown Fig 1(b).



(a) (b)

Fig. 1. (a) P-map of the differences in the hippocampus between NC and AD. Left: inferior view, right: superior view. (b) Iso-contour on p-maps in the left (top) and right (below) hippocampus. The views from left to right are lateral, superior, medial and inferior.

Table 1. Mean (SD) annualised atrophy rates, number of landmarks after thresholding (M), Cohen's d as effect size and the area under the receiver operating curve (AUC) using different p-value thresholds. * denotes statistical differences at $p < 0.05$ in AUC between 'no threshold' and 'threshold at the given p-value'.

	Annualised atrophy rate (%)			
	Converters (N=86)	Stable (N=128)	**Cohen's d**	**AUC**
No threshold (M=8196)	4.14 (3.64)	3.14 (3.08)	**0.30**	**0.59**
$p < 0.05$ (M=3174)	2.50 (1.94)	1.69 (1.77)	**0.44**	**0.63***
$p < 0.01$ (M=1650)	1.70 (1.30)	1.06 (1.25)	**0.51**	**0.65***
$p < 0.005$ (M=1107)	1.29 (1.00)	0.74 (0.96)	**0.57**	**0.66***
$p < 0.001$ (M=474)	0.74 (0.59)	0.40 (0.54)	**0.61**	**0.67***
$p < 0.0005$ (M=301)	0.53 (0.48)	0.27 (0.42)	**0.58**	**0.66***
$p < 0.0001$ (M=35)	0.08 (0.18)	0.04 (0.16)	**0.26**	**0.55**

Table 1 shows the mean (SD) of the annualised hippocampal atrophy rates, number of landmarks after thresholding, Cohen's d as effect sizes and the area under the receiver operating curve as a measure of overall classification performance (Fig. 2) using different p-value thresholds. Both Cohen's d and AUC (i.e. the power to classify MCI converters and stable) were higher when using ROIs generated using $p < 0.05$ to $p < 0.001$ compared to using the full hippocampus. Cohen's d and AUC reached a maximum when using ROIs generated by $p < 0.001$ to calculate the atrophy rate.

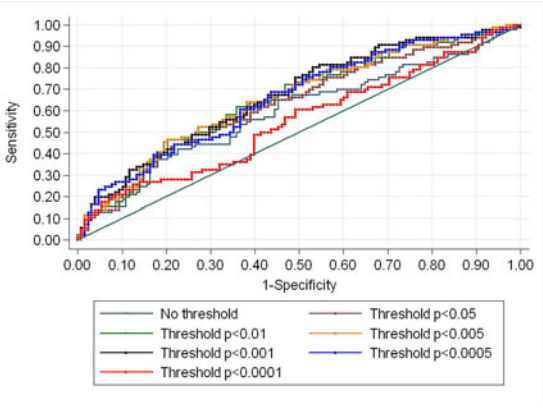


Fig. 2. Receiver operating curve of different p-value thresholds

4 Conclusions and Discussions

We have demonstrated that the power to classify MCI converters and stable subjects using hippocampal atrophy rates can be increased by using local ROIs within the hippocampus identified using SSM on a training set of NC and AD. An improvement in the early diagnosis of AD (at an MCI stage) is important for clinical and research purposes.

In this study, we have used data from NC and AD, which is independent of the MCI data, to identify ROIs that can increase the power to classify MCI converters and stable subjects. And we have found that the regions identified by SSM on the lateral and medial sides of the right hippocampus include, but are not limited to, CA1 and subiculum. Differences in regions that are near CA1 have been consistently reported between NC and AD [15,16]. The subiculum has also been shown to be different between NC and AD [15].

We found that thresholding at $p < 0.001$ provided the best classification power to separate MCI converters and stable subjects. Even though the atrophy rates using the $p < 0.001$ threshold proved to be more discriminant between MCI converters and stable subjects, they were about 17 times smaller than the ones obtained using the full hippocampus. This rather large difference in atrophy rate reflects the characteristics of the SSMs which were built using a similarity transform. The similarity transform removed global shape variations due to scaling, which is related to global atrophy. This technique is therefore more adapted to capture local regions which undergo larger change in shape than the rest of the hippocampus and allows better discrimination between the groups. It should be noted that longer follow-up is required since some of the stable subjects may convert to AD in the future.

As p-values are highly dependent on the number of samples in the training set, a different 'best p-value' is likely to be found when using a different number of samples. In future, we plan to use a effect size map instead of p-map to identify ROIs.

We have used the hippocampal segmentations from SNT to generate SSM because they are readily available from the ADNI website. However, there are different hippocampal segmentation protocols which produce different shapes of the hippocampus [17]. Our method may be applied to hippocampal segmentations from different protocols and methods [18] to determine which one has the most power to classify MCI converters and stable subjects.

Our method involves two image registration processes – inter-subject registrations between the baseline images in SSM and intra-subject registrations between baseline and repeat images in BSI. The two registration processes should be consistent, so that the statistically different ROIs identified in SSM will correspond to the atrophic regions between the registered baseline and repeat images in BSI. This is especially important for structures that are symmetrical in shape.

In conclusion, our method provides a framework to increase the statistical power of patient classification using the quantification of local atrophy rate in ROIs identified using shape differences between diagnostic groups.

References

1. Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E.: Mild cognitive impairment: clinical characterization and outcome. Arch. Neurol. 56(3), 303–308 (1999)
2. Apostolova, L.G., Dutton, R.A., Dinov, I.D., Hayashi, K.M., Toga, A.W., Cummings, J.L., Thompson, P.M.: Conversion of mild cognitive impairment to Alzheimer disease predicted by hippocampal atrophy maps. Arch. Neurol. 63(5), 693–699 (2006)
3. Chtelat, G., Fouquet, M., Kalpouzos, G., Denghien, I., la Sayette, V.D., Viader, F., Mzenge, F., Landeau, B., Baron, J.C., Eustache, F., Desgranges, B.: Three-dimensional surface mapping of hippocampal atrophy progression from MCI to AD and over normal aging as assessed using voxel-based morphometry. Neuropsychologia 46(6), 1721–1731 (2008)
4. Chupin, M., Garardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehericy, S., Benali, H., Garnero, L., Colliot, O.: Alzheimer's Disease Neuroimaging Initiative: Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19(6), 579–587 (2009)
5. Thompson, P.M., Hayashi, K.M., Zubicaray, G.I.D., 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 disease. Neuroimage 22(4), 1754–1766 (2004)
6. Hua, X., Lee, S., Yanovsky, I., Leow, A.D., Chou, Y.Y., Ho, A.J., Gutman, B., Toga, A.W., Jack, C.R., Bernstein, M.A., Reiman, E.M., Harvey, D.J., Kornak, J., Schuff, N., Alexander, G.E., Weiner, M.W., Thompson, P.M.: Alzheimer's Disease Neuroimaging Initiative: Optimizing power to track brain degeneration in Alzheimer's disease and mild cognitive impairment with tensor-based morphometry: an ADNI study of 515 subjects. Neuroimage 48(4), 668–681 (2009)

7. Morra, J.H., Tu, Z., Apostolova, L.G., Green, A.E., Avedissian, C., Madsen, S.K., Parikshak, N., Hua, X., Toga, A.W., Jack, C.R., Schuff, N., Weiner, M.W., Thompson, P.M.: Alzheimer's Disease Neuroimaging Initiative: Automated 3D mapping of hippocampal atrophy and its clinical correlates in 400 subjects with Alzheimer's disease, mild cognitive impairment, and elderly controls. Hum. Brain Mapp. 30(9), 2766–2788 (2009)
8. Haller, J.W., Banerjee, A., Christensen, G.E., Gado, M., Joshi, S., Miller, M.I., Sheline, Y., Vannier, M.W., Csernansky, J.G.: Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. Radiology 202(2), 504–510 (1997)
9. Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Shaw, L.M., Trojanowski, J.Q., Thompson, P.M., Jack, C.R., Weiner, M.W., Initiative, A.D.N.: MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 132(Pt 4), 1067–1077 (2009)
10. Davies, R.H., Twining, C.J., Taylor, C.: Groupwise surface correspondence by optimization: representation and regularization. Med. Image Anal. 12(6), 787–796 (2008)
11. Freeborough, P., Fox, N.: The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Transactions in Medical Imaging 16(5), 623–629 (1997)
12. Barnes, J., Foster, J., Boyes, R.G., Pepple, T., Moore, E.K., Schott, J.M., Frost, C., Scahill, R.I., Fox, N.C.: A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. Neuroimage 40(4), 1655–1671 (2008)
13. Hobbs, N.Z., Henley, S.M.D., Wild, E.J., Leung, K.K., Frost, C., Barker, R.A., Scahill, R.I., Barnes, J., Tabrizi, S.J., Fox, N.C.: Automated quantification of caudate atrophy by local registration of serial MRI: evaluation and application in Huntington's disease. Neuroimage 47(4), 1659–1665 (2009)
14. Leung, K.K., Clarkson, M.J., Bartlett, J.W., Clegg, S., Jack, C.R., Weiner, M.W., Fox, N.C., Ourselin, S.: Alzheimer's Disease Neuroimaging Initiative: Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: tissue-specific intensity normalization and parameter selection. Neuroimage 50(2), 516–523 (2010)
15. Wang, L., Swank, J.S., Glick, I.E., Gado, M.H., Miller, M.I., Morris, J.C., Csernansky, J.G.: Changes in hippocampal volume and shape across time distinguish dementia of the Alzheimer type from healthy aging. Neuroimage 20(2), 667–682 (2003)
16. Mueller, S.G., Weiner, M.W.: Selective effect of age, Apo e4, and Alzheimer's disease on hippocampal subfields. Hippocampus 19(6), 558–564 (2009)
17. Konrad, C., Ukas, T., Nebel, C., Arolt, V., Toga, A.W., Narr, K.L.: Defining the human hippocampus in cerebral magnetic resonance images–an overview of current segmentation protocols. Neuroimage 47(4), 1185–1195 (2009)
18. Leung, K.K., Barnes, J., Ridgway, G.R., Bartlett, J.W., Clarkson, M.J., Macdonald, K., Schuff, N., Fox, N.C., Ourselin, S.: Alzheimer's Disease Neuroimaging Initiative: Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and Alzheimer's disease. Neuroimage 51(4), 1345–1359 (2010)

Consistent 4D Cortical Thickness Measurement for Longitudinal Neuroimaging Study

Yang Li[1], Yaping Wang[2,1], Zhong Xue[3], Feng Shi[1], Weili Lin[1], Dinggang Shen[1], and The Alzheimer's Disease Neuroimaging Initiative*

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill, USA
[2] Department of Automation, Northwestern Polytechnical University, Xi'an, Shaanxi Province, P.R. China
[3] Methodist Center for Biotechnology and Informatics, The Methodist Hospital Research Institute, Weill Cornell Medical College, and Department of Radiology, The Methodist Hospital, USA

Abstract. Accurate and reliable method for measuring the thickness of human cerebral cortex provides powerful tool for diagnosing and studying of a variety of neuro-degenerative and psychiatric disorders. In these studies, capturing the *subtle* longitudinal changes of cortical thickness during pathological or physiological development is of great importance. For this purpose, in this paper, we propose a 4D cortical thickness measuring method. Different from the existing temporal-independent methods, our method fully utilizes the 4D information given by temporal serial images. Therefore, it is much more resistant to noises from the imaging and pre-processing steps. The experiments on longitudinal image datasets from the Alzheimer's Disease Neuroimaging Initiative (ADNI) show that our method significantly improves the longitudinal stability, *i.e.* temporal consistency, in cortical thickness measurement, which is crucial for longitudinal study. Power analysis of the correlation between cortical thickness and Mini-Mental-Status-Examination (MMSE) score demonstrated that our method generates statistically more significant results when comparing with the 3D temporal-independent thickness measuring methods.

1 Introduction

Many recent anatomical magnetic resonance image (MRI) studies on the human brain have been focused on the cerebral cortex thickness analysis, because longitudinal variations in cortical thickness are found closely correlated to either

* Data used in the preparation of this article were obtained from the Alzheimers Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. ADNI investigators include (complete listing available at www.loni.ucla.edu/ADNI/Collaboration/ADNI_Authorship_list.pdf).

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 133–142, 2010.

pathological (*e.g.* Alzheimer's disease) or physiological (*e.g.* normal aging) development of brains. Therefore, an accurate cortical thickness measuring method with longitudinal consistency and stability, which can detect and monitor the developmental changes of cortical thickness, is highly desirable. Many cortical thickness methods have been previously proposed. They can be broadly categorized as explicit surface based, implicit surface based and probabilistic segmentation based. In the explicit surface-based methods, after the inner (WM/GM interface) and outer (GM/CSF interface) surfaces are extracted by deformable surface models (which incorporate the smoothness constraints), the thickness is defined as the distance between a pair of points from each of the two surfaces. The correspondence between the two points in the pair is found either by deformable mapping of the inner surface to the outer surface [1,2], nearest point [3] or surface normal [4]. The disadvantages of using explicit surface are the extra computational cost and errors generated by the surface construction. In contrast, in implicit surface-based approaches, after segmentation of brain tissues, no surface mesh is explicitly constructed to represent the WM/GM and GM/CSF boundary. This makes the algorithms in this category more computationally efficient. The PDE-based [5] method is one of the representative approaches in this category [6,7,8]. In this method, Laplace's equation is solved in the GM region with certain boundary conditions (by setting different constant potentials on the two boundaries). The nested sub-layers of cortex is thus revealed by the resultant iso-potentials. The cortical thickness is then defined at each point as the length of the streamline along the gradient of the defined potential field. Some of the above methods are based on a hard segmentation of brain tissues. The disadvantage of using hard segmentation is the losing of sub-voxel information, which makes the algorithms very sensitive to the segmentation errors. To overcome this limitation, methods which measure the thickness on probabilistic segmentation of GM are proposed. Diffeomorphic registration of the probabilistic segmentation image is used in [9] to find a one-to-one correspondence of point pairs, between which cortical thickness is defined. Similarly, in [10], thickness is defined as the minimum line integral across the probablistic GM segmentation.

The above existing methods can be considered as 3D thickness measuring approaches, because they are designed to measure the thickness temporal-independently and do not take into account the temporal correlation. In order to improve the measuring accuracy and stability in longitudinal thickness studies, in this paper, we aim to devise a 4D thickness measuring method which is capable of fully utilizing the temporal information provided by longitudinal image dataset. The reason why 4D thickness measurement is important and necessary is that, measuring cortical thickness from MR images is affected by many artifacts and noises, such as intensity inhomogeneity, partial volume (PV) effect and segmentation errors. Comparing the thickness of cortical structures ($1.2 \sim 4.5mm$ [11]) to the the resolution of MR images ($\approx 1mm$), the errors introduced in the measuring process are considerably large. Therefore, if the thickness changes are evaluated as the difference between two temporal-independent 3D measurements, these errors will be amplified and result in jittery longitudinal

measurements. Another fact makes the longitudinal study even more difficult is that the expected change in GM thickness during the early stages of some neurological disorders, *e.g.* Alzheimer's disease, has been shown to be less than $1mm$ in most brain regions [12,13]. Since the cortical structure are only a few voxels thick in the images, sub-voxel accuracy is required to detect the subtle longitudinal thickness changes. In this situation, incorporating the information from other time-points as constraints to improve the accuracy and robustness of thickness measurement becomes very important. Currently, such temporal constraints are introduced by applying some sort of regression over the independently estimated 3D measurements. The problem of this regression-based method is that it imposes overly restrictive constraints due to the limitation of the pre-assumed regression model (usually linear), without taking into account the temporal correlation. This prompts us to incorporate the temporal constraints directly into the thickness measurement process. In this paper, after getting the 4D segmentation results of the longitudinal input images using a 4D segmentation method [14], we propose to measure the thickness on the aligned GM probability maps of different time-points in a common stereotaxic space. In this way, information from all time-points can be easily incorporated.

2 Methods

2.1 Cortical Thickness Measurement by Minimum Line Integral

In [10], a 3D thickness measurement is defined as the minimum line integral on the probabilistic segmentation of GM. As shown in Fig. 1, the thickness on each voxel in the GM (denoted by red dot) is defined as the minimum line integral (denoted as yellow arrow) of the probability map of the GM (the underneath image) over all possible directions (denoted as a group of green arrows) passing through that voxel. Mathematically, this method can be expressed as:

$$T(\mathbf{x}) = \min_{l \in L_{\mathbf{x}}} \int_l P(\mathbf{x}) \mathrm{d}l \tag{1}$$

where $T(\mathbf{x})$ is the measured thickness of cortex at voxel $\mathbf{x} \in \mathbb{R}^3$. $P(\mathbf{x}) \in [0\ 1]$ is the probability of the point $\mathbf{x}$ belonging to the GM. $L_{\mathbf{x}}$ is the set of all possible lines in three-dimensional space passing through $\mathbf{x}$. In implementation, $L_{\mathbf{x}}$ is defined by evenly sampling on the unit sphere.

2.2 4D Cortical Thickness Measurement

As Fig. 2 shows, our 4D thickness measurement method consists of five steps. Without loss of generality, we use a longitudinal image dataset with one baseline scan and $N-1$ follow-up scans to explain each step in the pipeline.

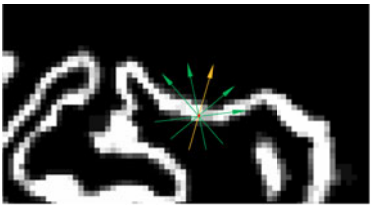


Fig. 1. Cortical thickness measurement by minimal line integral

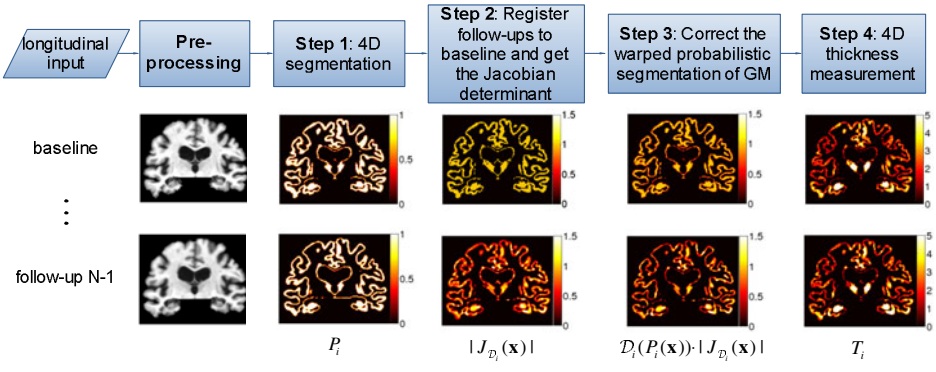


Fig. 2. 4D cortical thickness measurement

Pre-processing: Intensity inhomogeneity is firstly corrected [15]. Next, the $N-1$ follow-up images are rigidly (6 DOFs) registered to the baseline images, because the images are from the same subject with only translational or rotational misalignments. After that, the skull and cerebellum are removed [16] in baseline. Finally, by applying the resultant mask in baseline onto the aligned follow-up scans, we get consistent skull and cerebellum removing results in all follow-up scans.

Step 1: Input the pre-processed images into a 4D segmentation algorithm (CLASSIC [14]), to acquire the segmentation of GM with higher accuracy and longitudinal consistency. The probabilistic 4D segmentations of GM are denoted as P_i, $i = 1, \cdots, N$.

Step 2: Register each follow-up to baseline using diffeomorphic demons registration [17]. The resultant deformation fields and warped probabilistic segmentation of GM are denoted as $\mathcal{D}_i$ and $\mathcal{D}_i(P_i)$, respectively. To quantify the expansion and contraction caused by transformation $\mathcal{D}_i$, on each voxel, the Jacobian determinant $|J_{\mathcal{D}_i}(\mathbf{x})|$ are calculated (as shown in step 2 of Fig. 2). This map will be used as a scaling factor to correct the warped GM probability map $\mathcal{D}_i(P_i)$ by preserving the probabilistic volume. The reason to impose the

diffeomorphic constraints in the inter-timepoint registration is to seek a minimal deforming path. This property makes the deformation along the radial direction (the direction thickness changes) on the cortex mantle. In [9], this property of diffeomorphic registration was also used to find the corresponding point pairs in thickness measurement.

Step 3: In order to preserve the probabilistic volume, the warped GM probability map $\mathcal{D}_i(P_i)$ is corrected by multiplying with the scaling factor:

$$P_i'(\mathbf{x}) = \mathcal{D}_i(P_i(\mathbf{x})) \cdot |J_{\mathcal{D}_i}(\mathbf{x})| \tag{2}$$

This local probabilistic volume preserving step is also known as *modulation* in voxel-based morphometry [18].

Step 4: In the order of the scan time, the warped and corrected GM probability map of each time-point can be represented as $P_1', \cdots, P_N'$. Since these maps are in a standardized stereotaxic space (baseline space) and the probabilistic volume is preserved, for each GM voxel in this space, the minimal line integral direction can be defined as:

$$l_{min} = \arg\min_{l \in L_{\mathbf{x}}} \frac{1}{N} \sum_{i=1}^{N} \int_l P_i'(\mathbf{x}) \mathrm{d}l \tag{3}$$

Different from the definition in 3D temporal-independent thickness measurement, l_{min} is the optimal thickness measuring direction not only for a single time-point, but for all the images in the longitudinal image dataset. This means information from different time-points is fully utilized as constraint and guidance in finding the optimal measuring direction, which is the key factor in thickness measuring algorithms. Therefore, l_{min} can be robustly estimated against the noises. The thickness at GM voxel $\mathbf{x}$ on the *i-th* time-point is then defined as the line integral on P_i' along l_{min}:

$$T_i(\mathbf{x}) = \int_{l_{min}} P_i'(\mathbf{x}) \mathrm{d}l \tag{4}$$

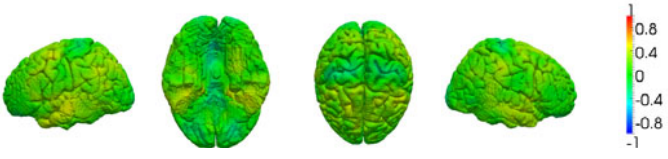
In this 4D measurement, we can make sure that the thicknesses to be compared in the longitudinal study are based on a common measuring direction. This will make all the thickness values resistant to outliers and much more comparable than the 3D measurement.

3 Experiments and Results

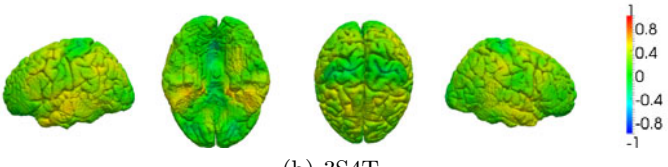
The validation of cortical thickness measuring algorithm has been a difficult problem, because no gold standard is available and can be used to evaluate a measurement. It is very difficult to manually measure the thickness in 3D images, due to the highly convoluted nature of cortex. Instead of the direct

validation, indirect validation method has been adopted to compare different thickness measuring methods [10,19]. This method is based on the fact that cortical thickness has close relationship with the psychological functions of brain. In some diseases, such as Alzheimer's disease, the decrease of thickness has been found highly correlated to the psychological disorder [20,21] which can be quantified by scores from some clinical examination, such as Mini-Mental-Status-Examination (MMSE) or Clinical Dementia Rating (CDR). Therefore, by comparing the correlations detected by different measuring method and the corresponding statistical significance, the accuracy and reliability of each method when applied in clinical studies can be evaluated [19].

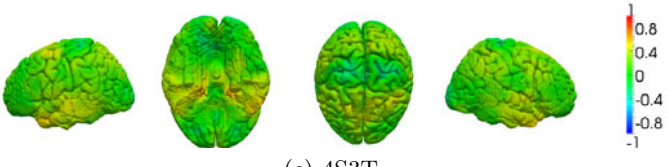
Data. Data used in the experiment were obtained from the public available ADNI database (www.loni.ucla.edu/ADNI). In our study, 40 patients with mild cognitive impairment (MCI) who later developed to probable Alzheimer's disease (AD) and, for comparison, 15 normal controls (NC) were selected. In MCI group, the average MMSE decline is 9.2 (within 2∼3 years), which indicates a substantial neuropsychological disorder has been developed. Therefore, the



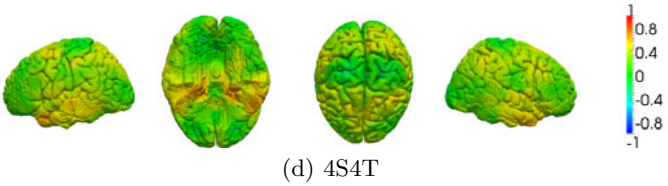
(a) 3S3T



(b) 3S4T



(c) 4S3T

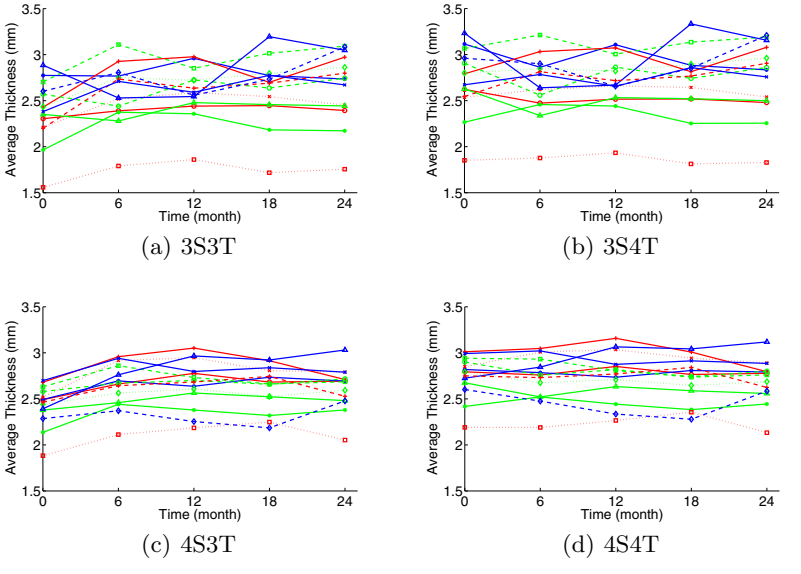


(d) 4S4T

Fig. 3. Average correlation between thickness and MMSE scores. From left to right: the left, inferior, superior and right views.

corresponding decrease of the cortical thickness is expected. In NC group, the average MMSE change is 1.3. Considering the possible MMSE assessment errors, subjects in this group can be regarded as neuropsychological healthy, and thus the cortical thickness is expected to keep stable (or slight decrease with normal aging).

Experiment design. Since the 4D processing is introduced at both the segmentation step and the thickness measuring step, totally four different thickness measuring pipelines are compared in order to trace the source of the possible observed improvements. The four different combinations are: 3D segmentation and 3D thickness measurement (3S3T), 3D segmentation and 4D thickness measurement (3S4T), 4D segmentation and 3D thickness measurement (4S3T) and 4D segmentation and 4D thickness measurement (4S4T). After the 4 different thickness values are measured, in order to conduct voxel-wise group analysis, each subject's thickness maps are mapped onto the template space.



(a) 3S3T (b) 3S4T

(c) 4S3T (d) 4S4T

Fig. 4. Longitudinal change of average thickness of 15 subjects from NC group. (Each curve represents a subject).

Results on MCI group. Before the voxel-wise correlation analysis can be conducted in the template space, the mapped thickness is first smoothed using full-width-at-half-maximum (FWHM) Gaussian filter ($\sigma = 8mm$) in order to suppress possible registration errors and inter-subject structure variations. After that, Pearson's correlation between thickness and MMSE score is calculated voxel-wise within each subject. The average correlation within the MCI

Table 1. Longitudinal stabilities of different thickness measuring methods

	3S3T	3S4T	4S3T	4S4T
regression residual ($mean \pm std$, mm)	0.161±0.067	0.121±0.073	0.117±0.052	0.081±0.040

group are then computed by transforming the correlation coefficients to Fisher's z-value and transforming back. The resultant average correlations from the four methods are summarized in Fig. 3. As we can see, when more 4D components are added into the pipeline, the higher correlation can be detected. The average correlation detected increases in the order $3S3T < 3S4T$ and $4S3T < 4S4T$. Among all the four methods, the fully 4D method (4S4T) gives the highest correlation. This shows that both 4D segmentation and 4D thickness measurement can improve the accuracy and consistency for longitudinal thickness analysis. Superior/Mid Temporal Pole, Entorhinal Cortex and Middle/Inferior Temporal Gyrus are the five ROIs, in which the highest correlation are detected. These findings are consistent with those from [20,21].

Results on NC group. In the experiment on NC group, the stability (robustness) of different methods are compared in the situation that the change of thickness is very slight. For each subject, the average whole cortical thickness is computed at every time-point. This longitudinal average thickness change of the 15 NC subjects are shown in Fig. 4. As we can see, 3D methods generate jittery changes which indicates the lack of longitudinal stability and consistency. In contrast, the proposed method is much more resistant to the noise and gives the most stable thickness measures. To quantitatively compare this stability, for each subject, a linear regression is performed on the longitudinal change curve and the fitting errors (residual) are computed. The results are summarized in Table 1. Our method gives the smallest mean residual (which can be viewed as the estimated mean error) and the smallest standard deviation.

4 Conclusion

We presented a 4D cortical thickness measuring framework. By applying the minimal line integral thickness measuring method on the aligned probability maps of GM from each time-point, we incorporate longitudinal information into the thickness measurement as temporal constraints. Experiments on clinical images from ADNI show that our method can detect much higher correlation between cortical thickness and MMSE scores with higher statistical significance. This indirectly indicates that our method is much more consistent and accurate in thickness measurement for longitudinal data.

References

1. Fischl, B., Dale, A.M.: Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. USA 97(20), 11050–11055 (2000)
2. Zeng, X., Staib, L., Schultz, R., Duncan, J.: Segmentation and measurement of the cortex from 3-D mr images using coupled-surfaces propagation. IEEE Trans. Med. Imag. 18(10), 927–937 (1999)
3. Miller, M., Massie, A., Ratnanather, J., Botteron, K., Csernansky, J.: Bayesian construction of geometrically based cortical thickness metrics. NeuroImage 12(6), 676–687 (2000)
4. Scott, M., Thacker, N.: Robust tissue boundary detection for cerebral cortical thickness estimation. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 878–885. Springer, Heidelberg (2005)
5. Jones, S., Buchbinder, B., Aharon, I.: Three-dimensional mapping of cortical thickness using laplace's equation. Hum. Brain Mapp. 11(1), 12–32 (2000)
6. Yezzi, A.J., Prince, J.L.: An Eulerian PDE approach for computing tissue thickness. IEEE Trans. Med. Imaging 22(10), 1332–1339 (2003)
7. Hutton, C., De Vita, E., Ashburner, J., Deichmann, R., Turner, R.: Voxel-based cortical thickness measurements in MRI. Neuroimage 40(4), 1701–1710 (2008)
8. Acosta, O., Bourgeat, P., Zuluaga, M.A., Fripp, J., Salvado, O., Ourselin, S.: Automated voxel-based 3D cortical thickness measurement in a combined Lagrangian-Eulerian PDE approach using partial volume maps. Med. Imag. Analysis 13(5), 730–743 (2009)
9. Das, S.R., Avants, B.B., Grossman, M., Gee, J.C.: Registration based cortical thickness measurement. NeuroImage 45(3), 867–879 (2009)
10. Aganj, I., Sapiro, G., Parikshak, N., Madsen, S.K., Thompson, P.M.: Measurement of cortical thickness from MRI by minimum line integrals on soft-classified tissue. Hum. Brain Mapp. 30(10), 3188–3199 (2009)
11. von Economo, C.: The Cytoarchitectonics of the Human Cerebral Cortex. Oxford Medical Publications, London (1929)
12. Lerch, J., Pruessner, J., Zijdenbos, A., Hampel, H., Teipel, S., Evans, A.: Focal decline of cortical thickness in Alzheimers disease identified by computational neuroanatomy. Cerebral Cortex 15(7), 995–1001 (2005)
13. Singh, V., Chertkow, H., Lerch, J., Evans, A., Dorr, A., Kabani, N.: Spatial patterns of cortical thinning in mild cognitive impairment and Alzheimers disease. Brain 129(11), 2885–2893 (2006)
14. Xue, Z., Shen, D., Davatzikos, C.: CLASSIC: Consistent longitudinal alignment and segmentation for serial image computing. Neuroimage 30(2), 388–399 (2006)
15. Sled, J.G., Zijdenbos, A.P., Evans, A.C.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imag. 17(1), 87–97 (1998)
16. Smith, S.: Fast robust automated brain extraction. Hum. Brain Mapp. 17(3), 143–155 (2002)
17. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1S1), S61–S72 (2009)
18. Ashburner, J., Friston, K.J.: Voxel-based morphometry–the methods. NeuroImage 11(6), 805–821 (2000)

19. Lerch, J.P., Evans, A.C.: Cortical thickness analysis examined through power analysis and a population simulation. NeuroImage 24(1), 163–173 (2005)
20. Desikan, R.S., Cabral, H.J., Hess, C.P., Dillon, W.P., Glastonbury, C.M., Weiner, M.W., Schmansky, N.J., Greve, D.N., Salat, D.H., Buckner, R.L., Fischl, B.: Automated MRI measures identify individuals with mild cognitive impairment and Alzheimer's disease. Brain 132(8), 2048–2057 (2009)
21. Holland, D., Brewer, J.B., Hagler, D.J., Fenema-Notestine, C., Dale, A.M.: Subregional neuroanatomical change as a biomarker for Alzheimer's disease. Proceedings of the National Academy of Sciences 106(49), 20954–20959 (2009)

Fiber-Centered Analysis of Brain Connectivities Using DTI and Resting State FMRI Data

Jinglei Lv[1], Lei Guo[1], Xintao Hu[1], Tuo Zhang[1], Kaiming Li[1,2], Degang Zhang[1,2], Jianfei Yang[1], and Tianming Liu[2]

[1] School of Automation, Northwestern Polytechnical University, Xi'an, China
[2] Department of Computer Science and Bioimaging Research Center, The University of Georgia, Athens, GA, USA

Abstract. Recently, inference of functional connectivity between brain regions using resting state fMRI (rsfMRI) data has attracted significant interests in the neuroscience community. This paper proposes a novel fiber-centered approach to study the functional connectivity between brain regions using high spatial resolution diffusion tensor imaging (DTI) and rsfMRI data. We measure the functional coherence of a fiber as the time series' correlation of two gray matter voxels that this fiber connects. The functional connectivity strength between two brain regions is defined as the average functional coherence of fibers connecting them. Our results demonstrate that: 1) The functional coherence of fibers is correlated with the brain regions they connect; 2) The functional connectivity between brain regions is correlated with structural connectivity. And these two patterns are consistent across subjects. These results may provide new insights into the brain's structural and functional architecture.

Keywords: functional network, structural network, rsfMRI, DTI, functional coherence.

1 Introduction

Recently analysis of functional connectivity between brain regions using resting state fMRI (rsfMRI) data has attracted significant interests in neuroscience community [1-3]. Functional connectivity between two brain regions is typically defined as the similarity of their fMRI time series. One of the major challenges in functional connectivity analysis of the human brain is the enormous size of possible combinations of brain regions. In the neuroimaging community, currently, there are two major streams of methodologies for functional connectivity study: ROI-based and clustering-based approaches. ROI-based approach is typically applied in applications that study limited number of brain regions. ROIs are either determined by manual placement of regions in the brain images or automatically determined by activation detection or other regional homogeneity analysis [4-6]. Clustering-based approaches are typically data-driven and they identify the brain networks that have more coherent fMRI time-series signals within each network [7, 8]. However, for high spatial resolution rsfMRI data, the computation time and memory consumption for volumetric clustering are increasing dramatically.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 143–150, 2010.

In this paper, we propose an intuitive fiber-centered approach to study the functional connectivity between brain regions using high spatial resolution diffusion tensor imaging (DTI) and rsfMRI data. Our premise is that axonal fibers obtained from DTI data are the structural substrates of functional connectivity between brain regions, and thus provide a natural anatomical localization for inference of functional connectivity. Therefore, we measure the correlation between rsfMRI time series of two gray matter voxels that this fiber connects to define the functional coherence of the fiber and the functional connectivity between the voxels it connect. We applied the above methodology to study the functional coherences of all fibers in the entire brain and to study the functional connectivity between major brain regions, which are parcellated via an atlas-based warping algorithm [9].

2 Method

2.1 Overview of the Method

The proposed computational pipeline is summarized in Figure 1. Firstly, we co-register rsfMRI, DTI and T1 images into the same space using the FSL FLIRT tool (*http://www.fmrib.ox.ac.uk/fsl/*). Then the whole brain is parcellated into regions by applying an atlas-based warping algorithms on the tissue map obtained from T1 image [9]. The anatomical labels are further mapped onto the DTI tissue map derived from [10]. After that, we group the tracked fibers from DTI data into clusters based on the brain regions that they connect. Finally, we extract the nearest gray matter voxels' rsfMRI signals for the two ends of each fiber, and perform functional coherence analysis, functional connectivity analysis and functional networks clustering based on these signals. This paper focuses on the last step.

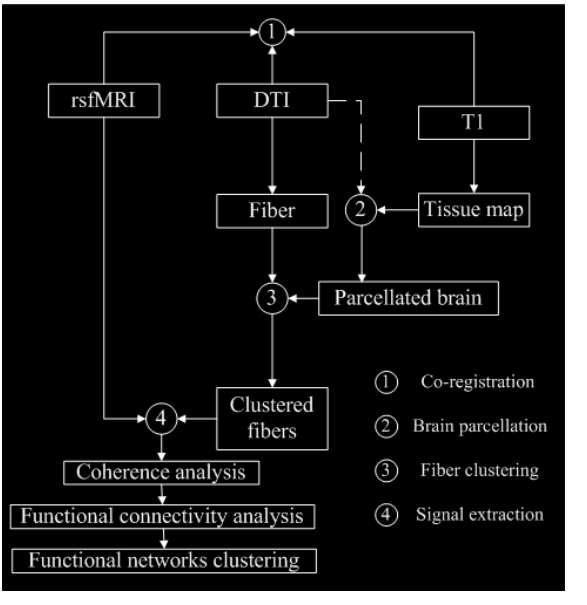


Fig. 1. The flowchart of the proposed computation pipeline for fiber-centered functional connectivity analysis.

2.2 Data Acquisition and Preprocessing

Eight volunteers were scanned using a 3T GE Signa MRI system. We acquired the resting state fMRI data with dimensionality 128×128×60×100, space resolution 2*mm*×2*mm*×2*mm*, TR 5*s*, TE 25*ms*, and flip angle 90 degrees. DTI data were acquired

using the same spatial resolution as the fMRI data. Parameters were TR 15.5*s* and TE 89.5*ms*, with 30 DWI gradient directions and 3 B0 volumes acquired.

Pre-processing of the rsfMRI data consists of brain skull removal, motion correction, spatial smoothing, temporal pre-whitening, slice time correction, global drift removal, and band pass filtering (0.01Hz~0.1Hz). The pre-processing of the DTI data consists of brain skull removal, motion correction, and eddy current correction. After the pre-processing, fiber tracking was performed using MEDINRIA (FA threshold: 0.2; minimum fiber length: 20; sampled by 4). Brain tissue segmentation was conducted on DTI data by our method in [10].

DTI space is used as the standard space from which to generate the GM segmentation and to report the brain network results on the cortical surface. Since fMRI and DTI sequences are both EPI sequences, their distortions tend to be similar and the misalignment between DTI and fMRI images is much less than that between T1 and fMRI images [11]. Co-registration between DTI and rsfMRI data is performed using the FSL FLIRT (http://www.fmrib.ox.ac.uk/fsl/).

2.3 Fiber Projection

There are a few reasons that cause the tracked fibers are not necessarily located on the cortex [12]. They include: 1) The FA values around the boundaries of gray matter and white matter are relatively low and the tractography procedure might stop before reaching the cortex. As a result, the tracked fibers will be within the cortical surface. 2) There is discrepancy in the brain tissue segmentation based on DTI data and the DTI tractography. In this case, the fiber could be either outside the cortex if the gray matter is over-segmented or inside the cortex if the gray matter is under-segmented [12]. Hence, in order to use the fiber connection information on the cortex, we project the fibers onto the cortical surface. If the end point of a fiber lies outside the cortex, we search along the fiber backwards the cortical surface. Otherwise, we extend the fiber towards the cortical surface. The searching process stops either when the fiber arrives at the surface or it exceeds a searching threshold. In very rare case when a fiber cannot reach the surface, we treat this fiber as an outlier and remove it. When a fiber are projected to the the surface, it has two HAMMER labels [9] of the surface patchs its ends pass through. Then, with the two labels, the fibers can be clustered into groups based on different combinations of start-end HAMMER label patterns, and these groups reflcet the structural connectivity between brain regions.

2.4 Functional Coherence Analysis for Fibers

As reported in the literature, the blood supply to the white matter is significantly lower than that of the cortex (less than one fourth) [13], and the BOLD contribution to the white matter is relatively low. So, the investigation of gray matter rsfMRI signals makes more sense. To ensure signals are extracted from the nearest gray matter voxel for each fiber end, after the co-registration of DTI and rsfMRI data, the tissue segmentation map of DTI image is used to guide the search for the nearest gray matter voxel. The searching method is similar to the one in Section 2.3, but this searching ends at the nearest voxel with gray matter label in the DTI-derived tissue segmentation image. Then, we extract the rsfMRI time series for two ends of each fiber, and define the functional coherence for each fiber as the absolute value of the Pearson's correlation coefficient of the two ends' rsfMRI signals.

2.5 Functional Connectivity between Brain Regions

In many previous studies of functional connectivity analysis, the fMRI signals are typically averaged within a brain region first, e.g., in a ROI (region of interest), and then the correlations between the averaged signals are computed as the functional connectivity. However, if the fMRI time series in a brain region are not coherent, the averaging process will smooth out real activity patterns. In this paper, the functional connectivity between two brain regions is defined as the average functional coherence of the fibers connecting them. Therefore the analysis of functional connectivity between two regions is decomposed into the analysis of functional connectivity of two structurally connected voxels. In this way, we avoided the risk of averaging inhomogeneous fMRI signals in a brain region.

To have anatomical meaningful regions of the brain, we first apply the HAMMER algorithm to parcellate the brain tissue obtained in Section 2.2 into 84 ROIs including both major cortical and sub-cortical structures [9, 14]. Then, for each hemisphere, we regroup these regions into 7 categories: the frontal lobe, the temporal lobe, the occipital lobe, the parietal lobe, subcortical region, cingulate region and the lateral ventricle. Therefore, we have 14 larger brain regions in addition. Using the method in Section 2.4, functional coherences of all fibers in the entire brain are calculated. Then we average the functional coherence of fibers connecting two regions to measure the functional connectivity strength between the two regions. So we have a symmetrical functional connectivity matrix C for all the 14 regions of the entire brain, and $c(i, j)$ measures the functional connectivity strength between region i and region j, $0 < c(i, j) < 1$.

2.6 Cluster Brain Regions Using Affinity Propagation Algorithm

The affinity propagation (AP) method [15] is a recently developed clustering method that can automatically determine the number of clusters in a population. It has been proved to have better performance than the traditional *K*-means clustering algorithm [15]. In general, The AP algorithm works by finding a set of exemplars in the data and assigning other data points to the exemplars. The AP algorithm approximates maximization of the sum of similarities to exemplars by recursively passing real-valued message along edge in a factor graph until a good set of exemplars and corresponding clusters emerges.

In this paper we use the AP algorithm to cluster the 84 regions into different functional networks, treating each region as a node in a network. There are two major steps in AP algorithm: similarity matrix computation and real-valued message passing, as summarized in following sections. Details on AP algorithm are referred to [15].

2.6.1 Definition of Similarity Matrix

The similarity matrix is obtained before the execution of the AP algorithm that simultaneously considers all data points as potential exemplars. We use the functional connectivity matrix as the similarity matrix. And the similarity between two regions is defined as $s(i,k) = -1/c(i,k)$, and for each i, $s(i, i)$ is set to 0. In this paper, we choose the maximum of the similarity multiplied by a constant value as self-similarity value p, and then we can obtain the expected clustering number by adjusting p value.

2.6.2 Real-Valued Message Passing Algorithm

Basically, there are two kinds of messages exchanged between data points, responsibility and availability, and each takes a different type of competition into account. The inputs are the pair-wise similarities and data point preferences.

In AP algorithm, the similarity $s(i,k)$ indicates how well the data point k is suited to be the exemplar for data point i. The preference $s(i, i)$ is defined such that points with high values are more likely to be selected as exemplars. The number of identified exemplars is affected by both the values of the input preference and the messaging-passing procedure. The responsibility $r(i, k)$, sent from data point i to a candidate exemplar point k, reflects the accumulated evidence of how well-suited point k is to serve as the exemplar for point i, taking into account other potential exemplars for point i. The self-responsibility $r(k, k)$ reflects accumulated evidence that point k is an exemplar based on its input preference tempered by how ill-suited it is to be assigned to another exemplar. The availability $a(i, k)$, sent from candidate exemplar point k to point i, indicates the accumulated evidence of how appropriate it would be for point i to choose point k as its exemplar, taking into account the support from other points that point k should be an exemplar. The availability $a(k, k)$, reflecting accumulated evidence that point k is an exemplar, is based on the positive responsibilities sent to candidate exemplar k from other points.

The update of responsibility lets all the candidate exemplars compete for ownership of a data point, whereas the update of availability collects evidence from data point as to whether each candidate exemplar would be a good exemplar. The messages of availability and responsibility are updated recursively. After convergence, availabilities and responsibilities are combined to identify exemplars. For point i, the value of k that maximizes $a(i,k)+r(i,k)$ either identifies point i as an exemplar if $k=i$, or identify data point that is the exemplar for point i. More details are referred to [15].

3 Results

3.1 Distribution of Fibers' Functional Coherences

We calculated the functional coherence of each fiber in the entire brain using the method mentioned in Section 2.4, and the statistic was done for all 8 subjects. We can conclude from Figure 2 that the functional coherence strength has a consistent probability distribution across subjects.

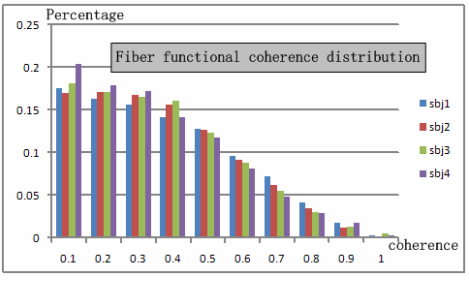


Fig. 2. Distributions of the fiber functional coherences of 4 randomly selected subjects

The functional coherence is also color coded for each fiber in the whole brain, as shown in Figure 3. The color of each fiber stands for the strength of functional coherence, and

the color bar is on the right. Figure 3 shows that the fiber functional coherence strength is correlated to the regions it connects. For instance, some groups of fibers connecting the temporal lobe have relatively lower functional coherences, while the inter-hemisphere fibers connecting the occipital lobes have relatively higher functional coherences. And this pattern is consistent across different brains, as shown in Figure 3 for 3 randomly selected subjects. This result demonstrates that the region-specific distribution of fibers' functional coherences is not random; it rather reflects brain's consistent functional architecture.

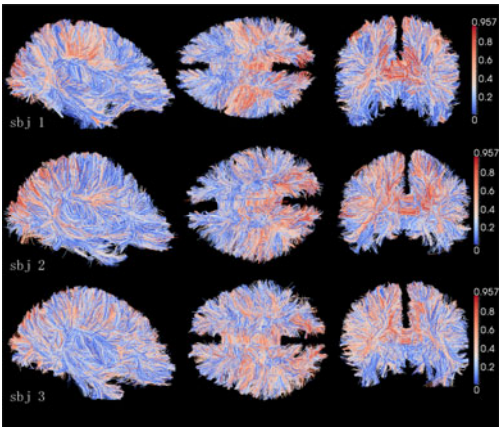


Fig. 3. Distributions of fibers' functional coherences over entire brain for 3 randomly selected subjects

3.2 Functional Connectivity Analysis

Based on the 14 regions described in section 2.5, we study the functional connectivity between these regions by constructing a connectivity matrix. We arrange the 14 regions in the order shown in Table.1. The distribution of the 14 regions is exactly symmetrical along the Y axis in Figure 4(a), which will help to view connectivity patterns.

The element (x, y) of the connectivity matrix is defined by the averaged functional coherence of all fibers connecting the region x and region y. Figure 4(a) shows the connectivity matrices for 4 randomly selected subjects. As we can see from the figure, 1) The connectivity matrix is relatively symmetrical along the dash dot line R; 2) The first and third quadrant is symmetrical along R, which means the functional connectivity between two hemispheres is relatively spatially symmetrical; 3) The second and fourth quadrant looks similar, demonstrating that the functional networks in each hemisphere are similar; 4) The connectivity pattern looks similar across different brains, but also with considerable variation.

Table 1. Region index

Region name	Region ID	
	Left	Right
Frontal lobe	1	14
Temporal lobe	2	13
Occipital lobe	3	12
Parietal lobe	4	11
Subcortical region	5	10
Cingulate region	6	9
Lateral ventricle	7	8

Similarly, based on the brain parcellation in step 2 (Figure 1), we performed functional connectivity analysis for 84 brain regions. The 84 regions are also arranged in a symmetrical order. In other words, the 14 regions are decomposed into smaller parts without changing the Region ID, and the smaller parts are also arranged to be symmetrical along the Y axis. The connectivity matrices are visualized in Figure 4(b). The connectivity matrices for the 84 regions further strongly support the conclusions drawn from 14 regions analysis above. However, it reflects more discrepancy between

two hemispheres, which is in consistency with the asymmetric brain. It is apparent that the connectivity pattern across subjects is quite reproducible. Also, the comparison of 14 and 84 regions analysis supports the idea that more segregated brain regions may have more functional differentiation.

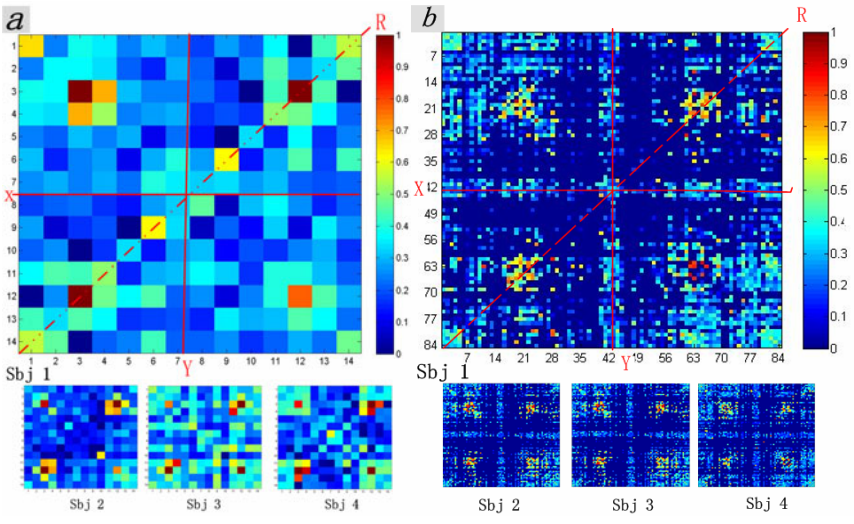


Fig. 4. (a) Functional connectivity matrices of 14 regions for 4 randomly selected subjects. (b) Functional connectivity matrices of 84 regions for 4 randomly selected subjects.

3.3 Region Clustering Using AP Algorithm

As we discussed in the previous section, the functional connectivity matrices for 84 regions show a stable pattern across subjects, which makes the clustering via AP algorithm feasible. We adjust the self-similarity value p in AP algorithm to obtain a stable clustering for each subject, and the cluster number for each case ranges from 20 to 24. The clustering result is mapped to cortical surface, and for each subject each color represents a sub-network, as shown in Figure 5 for 4 randomly selected subjects. Note that the colors are just randomly selected to show sub-network boundaries for each subject. They may vary across subjects for the same functional sub-network. The clustering result looks reasonable by visual inspection. As we can see in the top of Figure 5, the visual cortical regions of two hemispheres are clustered

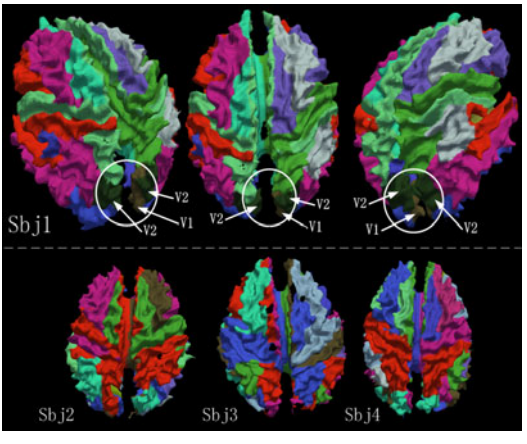


Fig. 5. Clustering of the 84 regions into sub-networks for randomly selected 4 regions

into meaningful regions, e.g., the V1 and V2 regions are clearly separated. This cluster is relatively consistent across subjects. We can see that the distributions of some sub-networks are symmetrical in two hemispheres. Some sub-networks contain regions across hemispheres. These clustering results support the view that the anatomical structural network is the basis of functional networks.

4 Discussion and Conclusion

In this paper, we present a novel fiber-centered approach for functional connectivity analysis in the entire brain. The advantages of this methodology include its intuitiveness and anatomical meaning. The analysis of functional connectivity between two brain regions is converted into the analysis of functional coherences of the fibers connecting these two regions. Our analysis on normal brains has shown meaningful results. In the future, we will cluster fibers using the functional coherences and their spatial distribution features. With the clustered fibers, we can potentially parcellate the white matter and gray matter into functionally homogenous regions automatically.

References

1. Fox, M.D., Raichle, M.E.: Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nat. Rev. Neurosci. 8(9), 700–711 (2007)
2. Damoiseaux, J.S., et al.: Consistent resting-state networks across healthy subjects. PNAS, 13848–13853 (2006)
3. Honey, C.J., et al.: Predicting human resting-state functional connectivity from structural connectivity. PNAS 106(6), 2035–2040 (2009)
4. Biswal, B., et al.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn. Reson. Med. 34(4), 537–541 (1995)
5. Cordes, D., et al.: Mapping functionally related regions of brain with functional connectivity MR imaging. AJNR Am. J. Neuroradiol. 21(9), 1636–1644 (2000)
6. Biswal, B., et al.: Simultaneous assessment of flow and BOLD signals in resting-state functional connectivity maps. NMR Biomed. 10(4-5), 165–170 (1997)
7. De Luca, M., et al.: fMRI resting state networks define distinct modes of long-distance interactions in the human brain. Neuroimage 29(4), 1359–1367 (2006)
8. van den Heuvel, M., et al.: Normalized cut group clustering of resting-state FMRI data. PLoS One, 3(4), e2001 (2008)
9. Shen, D., et al.: HAMMER: hierarchical attribute matching mechanism for elastic registration. IEEE Trans. Med. Imaging 21(11), 1421–1439 (2002)
10. Liu, T., et al.: Brain Tissue Segmentation Based on DTI Data. NeuroImage 38(1), 114–123 (2007)
11. Li, K., et al.: Cortical surface based identification of brain networks using high spatial resolution resting state FMRI data. In: IEEE International Symposium on Biomedical Imaging (ISBI), Rotterdam, pp. 657–659 (2010)
12. Zhang, D., et al.: Automatic cortical surface parcellation based on fiber density information. In: IEEE International Symposium on Biomedical Imaging (ISBI), Rotterdam, pp. 1133–1136 (2010)
13. Mezer, A., et al.: Cluster analysis of resting-state fMRI time series. NeuroImage 45, 1117–1125 (2009)
14. Liu, T., et al.: Deformable Registration of Cortical Structures via Hybrid Volumetric and Surface Warping. NeuroImage 22(4), 1790–1801 (2004)
15. Frey, B.J., Dueck, D.: Clustering by Passing Messages between Data Points. Science 315, 972–976 (2007)

A Generative Model for Brain Tumor Segmentation in Multi-Modal Images

Bjoern H. Menze[1,2], Koen Van Leemput[1,3,4], Danial Lashkari[1], Marc-André Weber[5], Nicholas Ayache[2], and Polina Golland[1]

[1] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, USA
[2] Asclepios Research Project, INRIA Sophia-Antipolis, France
[3] Radiology, Massachusetts General Hospital, Harvard Medical School, USA
[4] Information and Computer Science, Aalto University, Finland
[5] Diagnostic Radiology, Heidelberg University Hospital, Germany

Abstract. We introduce a generative probabilistic model for segmentation of tumors in multi-dimensional images. The model allows for different tumor boundaries in each channel, reflecting difference in tumor appearance across modalities. We augment a probabilistic atlas of healthy tissue priors with a latent atlas of the lesion and derive the estimation algorithm to extract tumor boundaries and the latent atlas from the image data. We present experiments on 25 glioma patient data sets, demonstrating significant improvement over the traditional multivariate tumor segmentation.

1 Introduction

Limited therapy options require a careful diagnostic for patients with brain tumors. A multitude of available brain imaging sequences gives rise to patient data sets that include multi-parametric, multi-modal, and multi-temporal volumes even in standard clinical settings. Quantitative analysis of a lesion in these data poses a challenging computational problem. In this paper, we present a fully automated method for channel-specific tumor segmentation in such multi-dimensional images.

Generative probabilistic models of spatial tissue distribution and appearance have enjoyed popularity for tissue classification as they exhibit good generalization to unseen images [1,2,3]. Encoding spatial prior knowledge for a lesion, however, is difficult. Tumors may be modeled as outliers relative to the expected shape [4,5] or image signal of healthy tissues [2,6]. In [2], for example, a criterion for detecting outliers is used to generate a tumor prior in a subsequent EM segmentation which is treating tumor as an additional tissue class. Alternatively, the spatial prior for the tumor can be derived from the appearance of tumor-specific bio-markers [7,8]. The tumor classification methods can be augmented with spatial regularization using a Markov Random Field prior [9] or a boundary finding step [2,10] to ensure spatial contiguity of the segmentation results.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 151–159, 2010.

Discriminative approaches directly learn the difference between the appearance of the lesion and other tissues and do not rely on spatial priors [11,12,13,14,15,16]. They do, however, often require substantial amounts of training data and typically come at the cost of manual interaction for initialization and postprocessing. Most require the imaging protocol to be exactly the same in the training set and in the novel images to be segmented. Discriminative approaches proposed for tumor segmentation may use specific anatomical [13], but also generic image features (e.g., wavelets [11]) as input to the classifier. A spatial regularization via boundary modeling [11,12,13] or Markov Random Fields [14,15,16] has proved useful when used with discriminative methods as well.

Both generative and discriminative models face significant challenges when applied to multi-modal data. Automatic discriminative approaches are limited to the image modalities of the training set and are sensitive to missing data. Generative models may generalize straightforwardly to multi-channel observations [8,7], but do not allow for modeling differences between the biological processes observed in different modalities. By assuming the same shape and extend of pathology in all modalities, the standard multi-channel segmentation may ignore much of the information potentially available in images. Examples include differences in tissue water (T_2, Flair-MRI), enhancement of contrast agents (post-Gadolinium T_1-MRI), diffusion (DTI, DCE-MRI), or relative concentrations of selected metabolites (MRSI). Delineating the tumor area in each of these modalities individually is highly preferred for subsequent quantitative analysis of tumor shape and evolution.

We present a tumor appearance model for such multi-dimensional sequences and derive an algorithm for a channel-specific segmentation of the tumor. The method shares information about the spatial location of the lesion among channels while making full use of the highly specific multi-modal signal of the healthy tissue classes for segmenting normal tissues in the brain. In addition to tissue types, the model includes a latent variable for each voxel encoding the probability of observing tumor at that voxel. We derive an estimation algorithm for this model that generalizes the standard atlas-based EM segmentation. In our experiment with 25 multi-modal image volumes, the proposed approach performs significantly better than the traditional multivariate tissue classification method that assumes a single tumor segmentation that is shared by all channels.

2 Generative Tumor Model

We use a generative modeling approach, in which we first build an explicit statistical model of image formation and subsequently use this model to derive a fully automatic segmentation algorithm. Fig. 1 illustrates our generative model.

We model the *normal state* of the healthy brain using a spatially varying probabilistic prior $\boldsymbol{\pi}_k$ for each of the K tissue classes (Fig. 1, blue). This prior

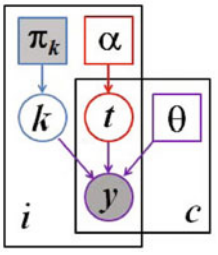


Fig. 1. Graphical model for the proposed segmentation approach. Voxels are indexed with i, the channels are indexed with c. The known prior π_k determines the label k of the normal, healthy tissue. The latent atlas α determines the channel-specific presence of tumor $\boldsymbol{t}$. Normal state k, tumor state $\boldsymbol{t}$, and intensity distribution parameters θ jointly determine the multi-modal image observations $\boldsymbol{y}$. Observed (known) quantities are shaded. The tumor segmentation aims to estimate $p(t_i^c|\boldsymbol{y})$, along with the segmentation of healthy tissue $p(k_i|\boldsymbol{y})$.

(atlas) is estimated from prior examples and is assumed to be known. At each voxel i, the atlas defines a multinomial distribution for the tissue label k_i:

$$p(k_i = k) = \pi_{ki}. \tag{1}$$

The normal state k_i is shared among all C channels at voxel i. In our experiments we assume $K = 3$, representing gray matter, white matter and cerebrospinal fluid (CSF).

We model the *tumor state* using a spatially varying "latent" probabilistic atlas $\boldsymbol{\alpha}$, similar to [10] (Fig. 1, red). At each voxel i, this atlas provides a scalar parameter α_i that defines the probability of observing tumor at that voxel. Parameter α_i is unknown and is estimated as part of the segmentation process. We define a latent tumor state $t_i^c \in \{0, 1\}$ that indicates the presence of tumor in channel c at voxel i and model it as a Bernoulli random variable with parameter α_i. We form a binary tumor state vector $\boldsymbol{t}_i = [t_i^1, \ldots, t_i^C]^T$ indicating the tumor presence for all c observations at voxel i, with probability

$$p(\boldsymbol{t}_i; \alpha_i) = \prod_c p(t_i^c; \alpha_i) = \prod_c \alpha_i^{t_i^c} \cdot (1 - \alpha_i)^{1 - t_i^c}. \tag{2}$$

Image observations y_i^c are generated by Gaussian intensity distributions for each of the K tissue classes and the C channels, with mean μ_k^c and variance v_k^c, respectively (Fig. 1, purple). In tumor tissue (i.e., if $t_i^c = 1$) the normal observations are replaced by intensities from another set of channel-specific Gaussian distributions with mean μ_{K+1}^c and variance v_{K+1}^c, representing the tumor class. Letting $\boldsymbol{\theta}$ denote the set of all mean and variance parameters, and $\boldsymbol{y}_i = [y_i^1, \ldots, y_i^C]^T$ denote the vector of the intensity observations at voxel i, we define the data likelihood:

$$\begin{aligned} p(\boldsymbol{y}_i|\boldsymbol{t}_i, k_i; \boldsymbol{\theta}) &= \prod_c p(y_i^c|t_i^c, k_i; \boldsymbol{\theta}) \\ &= \prod_c \left[\mathcal{N}(y_i^c \ ; \ \mu_{k_i}^c, v_{k_i}^c)^{1 - t_i^c} \cdot \mathcal{N}(y_i^c \ ; \ \mu_{K+1}^c, v_{K+1}^c)^{t_i^c} \right], \end{aligned} \tag{3}$$

where $\mathcal{N}(\cdot \ ; \ \mu, v)$ is the Gaussian distribution with mean μ and variance v.

Finally, the joint probability of the the latent atlas and the observed variables

$$p(\boldsymbol{y}_i, \boldsymbol{t}_i, k_i; \boldsymbol{\theta}, \alpha_i) = p(\boldsymbol{y}_i|\boldsymbol{t}_i, k_i; \boldsymbol{\theta}) \cdot p(\boldsymbol{t}_i; \alpha_i) \cdot p(k_i) \tag{4}$$

is the product of the components defined in Eqs. (1-3).

3 Maximum Likelihood Parameter Estimation

We seek Maximum Likelihood estimates of the model parameters $\{\boldsymbol{\theta}, \boldsymbol{\alpha}\}$:

$$\langle \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}} \rangle = \arg\max_{\langle \boldsymbol{\theta}, \boldsymbol{\alpha} \rangle} \; p(\boldsymbol{y}_1, \ldots, \boldsymbol{y}_N; \boldsymbol{\theta}, \boldsymbol{\alpha}) = \arg\max_{\langle \boldsymbol{\theta}, \boldsymbol{\alpha} \rangle} \; \prod_{i=1}^{N} p(\boldsymbol{y}_i; \boldsymbol{\theta}, \boldsymbol{\alpha}),$$

where N is the number of voxels in the volume and

$$p(\boldsymbol{y}_i; \boldsymbol{\theta}, \boldsymbol{\alpha}) = \sum_{\boldsymbol{t}_i} \sum_{k_i} p(\boldsymbol{y}_i, \boldsymbol{t}_i, k_i; \boldsymbol{\theta}, \boldsymbol{\alpha}).$$

Observing that evaluating the objective function involves summing over values of $\boldsymbol{t}_i$ and k_i in Eq. (4), we use Jensen's inequality to perform the optimization using an iterative, EM-style minorization technique [17]. Letting $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\alpha}}\}$ denote the current parameter estimates, we can compute the posterior probability of any of the 2^C tumor state vectors $\boldsymbol{t}_i$, writing out the components of Eq. (4):

$$q_i(\boldsymbol{t}_i) \;\triangleq\; p(\boldsymbol{t}_i | k_i, \boldsymbol{y}_i; \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\alpha}}) \;\propto\; \sum_{k_i} p(\boldsymbol{y}_i | \boldsymbol{t}_i, k_i; \tilde{\theta}) p(\boldsymbol{t}_i; \tilde{\alpha}_i) p(k_i), \tag{5}$$

and $\sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i) = 1$. Based only on the intensity channels that do *not* show tumor ($t_i^c = 0$), we also compute the posterior probability of tissue k at voxel i:

$$w_{ik}(\boldsymbol{t}_i) \;\triangleq\; p(k_i | \boldsymbol{t}_i, \boldsymbol{y}_i; \widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\alpha}}) \;\propto\; \pi_{ki} \prod_c \mathcal{N}(y_i^c \;;\; \tilde{\mu}_k^c, \tilde{v}_k^c)^{1-t_i^c},$$

and $\sum_k w_{ik}(\boldsymbol{t}_i) = 1$ for all $\boldsymbol{t}_i$. Using $q_i(\cdot)$ and $w_{ik}(\cdot)$, we arrive at closed-form update expressions that guarantee increasingly better estimates of the model parameters. The updates are intuitive: the latent tumor prior is an average of the corresponding posterior estimates

$$\tilde{\alpha}_i \leftarrow \sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i) \left(\frac{1}{C} \sum_c t_i^c \right)$$

and the intensity parameters are updated with the weighted statistics of the data for the healthy tissues ($k = 1, \ldots, K$)

$$\tilde{\mu}_k^c \leftarrow \frac{\sum_i \sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i)\, w_{ik}(\boldsymbol{t}_i)(1 - t_i^c)\, y_i^c}{\sum_i \sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i)\, w_{ik}(\boldsymbol{t}_i)(1 - t_i^c)}, \tilde{v}_k^c \leftarrow \frac{\sum_i \sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i)\, w_{ik}(\boldsymbol{t}_i)(1 - t_i^c)\, (y_i^c - \tilde{\mu}_k^c)^2}{\sum_i \sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i)\, w_{ik}(\boldsymbol{t}_i)(1 - t_i^c)}$$

and for the tumor class:

$$\tilde{\mu}_{K+1}^c \leftarrow \frac{\sum_i \sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i)\, t_i^c\, y_i^c}{\sum_i \sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i)\, t_i^c}, \quad \tilde{v}_{K+1}^c \leftarrow \frac{\sum_i \sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i)\, t_i^c\, (y_i^c - \tilde{\mu}_{K+1}^c)^2}{\sum_i \sum_{\boldsymbol{t}_i} q_i(\boldsymbol{t}_i)\, t_i^c}.$$

We iterate the estimation of the parameters $\{\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\alpha}}\}$ and the computation of the posterior probabilities $\{q_i(\cdot), w_{ik}(\cdot)\}$ until convergence.

4 Tumor Segmentation

Once we have an estimate of the model parameters $\{\widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}\}$, we can evaluate the probability that tumor is visible in channel c of voxel i by summing over all the configurations $\boldsymbol{t}_i$ for which $t_i^c = 1$:

$$p(t_i^c = 1|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) = \sum_{\boldsymbol{t}_i} t_i^c \; p(\boldsymbol{t}_i|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) = \sum_{\boldsymbol{t}_i} t_i^c \, q_i(\boldsymbol{t}_i). \tag{6}$$

We then assign channel c of voxel i to tumor if $p(t_i^c = 1|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) > 0.5$.

5 Extensions

To augment the generative model outlined above with further physiological knowledge, we derive and implement extensions considering the expected shape, multivariate signal and structural appearance of the tumor.

Little spatial context is used in the basic model, as we assume the tumor state $\boldsymbol{t}_i$ in each voxel to be independent from the state of other voxels (Eq. 6 and Eq. 3). It is only the atlas $\boldsymbol{\pi_k}$ that encourages smooth classification for the healthy tissue classes by imposing similar priors in neighboring voxels. To encourage a similar smoothness of the tumor labels, we extend the latent atlas α to include a Markov Random Field (MRF) prior:

$$p(\mathbf{t}_1, \ldots, \mathbf{t}_N; \beta, \boldsymbol{\alpha}) \propto \prod_c \prod_i \alpha_i^{t_i^c} (1-\alpha_i)^{1-t_i^c} \exp\left[-\frac{\beta}{2} \sum_{j \in \mathcal{N}_i} t_i^c (1-t_j^c) + t_j^c (1-t_i^c) \right].$$

Here, $\mathcal{N}_i$ denotes the set of the six nearest neighbors of voxel i, and β is a parameter governing how similar the tumor states tend to be in neighboring voxels. When $\beta = 0$, there is no interaction between voxels and the model reduces to the one described in Section 2. For $\beta \neq 0$, the posteriors $q_i(\boldsymbol{t}_i)$ are no longer given by Eq. (5), and their exact computation becomes infeasible. However, relaxing the MRF to a mean-field approximation [18] we derive an efficient approximate algorithm. We let

$$n_i^c = \sum_{j \in \mathcal{N}_i} \sum_{\boldsymbol{t}_j} t_j^c \, q_j(\boldsymbol{t}_j)$$

denote the currently estimated "soft" count of neighbors that show tumor in channel c. The mean-field approximation implies

$$p(\boldsymbol{t}_i|\alpha_i) \;\simeq\; \prod_c \gamma_i^{t_i^c} (1-\gamma_i)^{(1-t_i^c)}, \quad \text{where} \quad \gamma_i = \frac{\alpha_i}{\alpha_i + (1-\alpha_i) \exp\left[-\beta(2n_i^c - 6) \right]}$$

to replace the previously defined $p(\boldsymbol{t}_i|\alpha_i)$ in Eq. (4), leading to smoothed estimates of the tumor segmentations.

Moreover, we want to account for the non-homogeneity in the appearance of the tumor class, as gliomas show characteristic substructures such as active and

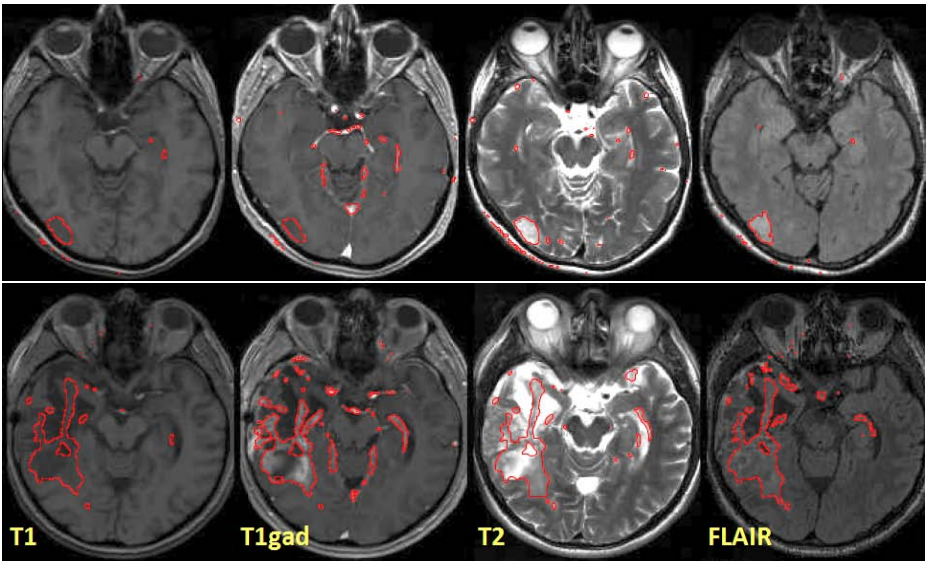


Fig. 2. Examples of channel-specific segmentation results for four different modalities, in two patients. The outlines of regions with $p(t_i^c = 1|\boldsymbol{y}_i; \widehat{\boldsymbol{\theta}}, \widehat{\boldsymbol{\alpha}}) > 0.5$ are shown in red. The proposed method localizes the tumor reliably in post-therapeutic images (below), where surgery has led to significant deviations from normalcy for healthy tissues.

necrotic areas and edema. We model this via a straightforward extensions of the tissue classes to include more than one class for tumor in a second modification to our approach. Finally, to consider higher-order interactions in the multivariate biological signal $\boldsymbol{y}_i$ of healthy tissue, we can relax the conditional independence of observations across channels by using multivariate Gaussians in the data likelihood in Eq. (3). We report tests of these three extensions in the next section.

6 Experiments

We evaluate our approach on a data set of 25 patients with glioma. The data set comprises T_1, T_2, FLAIR-, and post-Gadolinium T_1 MR images. Tumors were outlined by a rater in three planes intersecting with the tumor center. We register all images of a patient to the FLAIR volume by using affine registration and segment the volume into the three healthy and an outlier class using a freely available implementation of the EM segmentation with bias correction [1]. Outliers are defined as being more than three standard deviations away from the centroid of any of the three normal tissue classes.

We apply our algorithm to the bias field corrected volumes and initialize intensity parameters with values estimated in the initial segmentation. When using multivariate distributions $\mathcal{N}(\cdot; \mu; V)$,we initialize off-diagonal element in V to zero. When modeling the tumor class with multiple Gaussian densities we initialize the means of additional subclasses to random values. We use outliers in

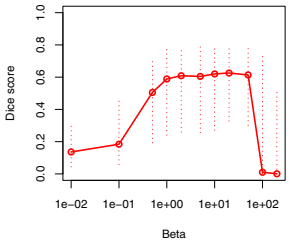


Fig. 3. Sensitivity to the MRF parameter β. Indicated are the median (solid line) and the interquartile ranges of the average Dice scores of all 25 data set. While some regularization is beneficial, the segmentation performance is relatively insensitive to the choice of the only model parameter β.

the initial segmentation to initialize the latent atlas $\boldsymbol{\alpha}$, setting α_i for pixels of the outlier class to 0.7 and otherwise to 0.3. We typically observe convergence after 10 to 15 steps. For comparison, we also implement an EM segmentation treating tumor as one of the tissue classes, with a weak MRF prior for spatial regularization, similar to [2]. We use the same data and initializations as above, but augment the atlas by a tumor prior obtained by smoothing the outlier class of the initial segmentation with a 3cm kernel. This alternative segmentation is applied to every single image volume individually in a first experiment, and to the multivariate features of the whole multi-modal volume in a second experiment. To evaluate the classification maps we calculate Dice scores for both methods [19].

Fig. 2 illustrates results for two different subjects. We note that the tumor boundaries change across different modalities and the proposed method captures this variation well, even in post-therapeutic images. The method produces few false positives which can be easily eliminated in a post-processing step. We evaluate the robustness and accuracy of our method in a series of experiments. First, we test the sensitivity of the performance to the MRF parameter β that governs the smoothness of the resulting segmentations. It is the only parameter to be adjusted in our model. We find the performance of the algorithm to be relatively stable for a wide range of β values (Fig. 3), irrespective of size, location or shape of the tumor (i.e., $\beta \in [1; 50]$). For simplicity, we set $\beta = 1$ in all further experiments. In the second experiment, we test different model options for normal

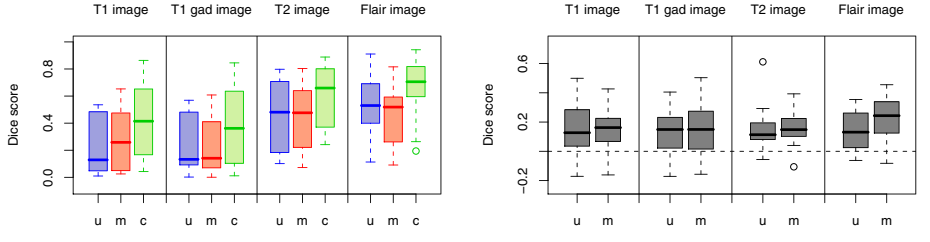


Fig. 4. Benefits of the channel-specific segmentation. Boxplots show median, quartiles and outliers for the Dice scores of all 25 subjects, for all four modalities. Our channel-wise segmentation (c, green) improves over both multiple univariate (u, blue) and multivariate (m, red) segmentation, both in the absolute terms (left) and with respect to patient-specific differences (right). The right figure shows $c-u$ and $c-m$.

tissue and tumor classes. We find little differences between approaches that use non-zero off-diagonal elements in the covariance matrix of the intensity likelihood and those that do not. Modeling tumor by three Gaussians improves the result for some cases, but leads to a somewhat lower performance on average. In a third experiment, we compare our approach to the two alternative EM segmentation methods (Fig. 4). Here, we find the proposed channel-specific segmentation to always perform significantly better ($p < 5 \cdot 10^{-4}$, paired Cox-Wilcoxon test); it improves the absolute value of the Dice score over all four modalities for nearly all data sets by 0.1 to 0.2 (Fig. 4, right).

7 Conclusions

We present a generative model for tumor appearance in multi-modal image volumes of the brain that provides channel-specific segmentations. We derive an estimation algorithm and demonstrate superior performance over standard multivariate EM segmentation. Unlike discriminative tumor segmentation methods, our model is applicable to any set of multi-modal image volumes, and is fully automatic. Further extensions of the model may consider structure of the tumor, or temporal evolution in longitudinal data sets.

Acknowledgements. This work was supported by the German Academy of Sciences Leopoldina (Fellowship Programme LPDS 2009-10), the Academy of Finland (133611), INRIA CompuTumor, NIH NIBIB NAMIC U54-EB005149, NIH NCRR NAC P41-RR13218, NIH NINDS R01-NS051826, NIH R01-NS052585, NIH R01-EB006758, NIH R01-EB009051, NIH P41-RR014075 and the NSF CAREER Award 0642971.

References

1. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE TMI 18, 897–908 (1999)
2. Prastawa, M., Bullitt, E., Ho, S., Gerig, G.: A brain tumor segmentation framework based on outlier detection. MedI. A 8, 275–283 (2004)
3. Pohl, K.M., Fisher, J., Levitt, J.J., Shenton, M.E., Kikinis, R., Grimson, W.E.L., Wells, W.M.: A unifying approach to registration, segmentation, and intensity correction. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 310–318. Springer, Heidelberg (2005)
4. Zacharaki, E.I., Shen, D., Davatzikos, C.: Orbit: A multiresolution framework for deformable registration of brain tumor images. IEEE TMI 27, 1003–1017 (2008)
5. Bach Cuadra, B., Pollo, C., Bardera, A., Cuisenaire, O., Thiran, J.P.: Atlas-based segmentation of pathological brain MR images using a model of lesion growth. IEEE TMI 23, 1301–1314 (2004)
6. Gering, D.T., Grimson, W.E.L., Kikinis, R.: Recognizing deviations from normalcy for brain tumor segmentation. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 388–395. Springer, Heidelberg (2002)

7. Moon, N., Bullitt, E., Van Leemput, K., Gerig, G.: Model-based brain and tumor segmentation. In: Proc ICPR, pp. 528–531 (2002)
8. Prastawa, M., Bullitt, E., Moon, N., Van Leemput, K., Gerig, G.: Automatic brain tumor segmentation by subject specific modification of atlas priors. Acad. Radiol. 10, 1341–1348 (2003)
9. Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., Suetens, P.: Automated segmentation of multiple sclerosis lesions by model outlier detection. IEEE TMI 20, 677–688 (2001)
10. Riklin-Raviv, T., Menze, B.H., Van Leemput, K., Stieltjes, B., Weber, M.A., Ayache, N., Wells III, W., Golland, P.: Joint segmentation via patient-specific latent anatomy model. In: Proc. MICCAI-PMMIA (2009)
11. Cobzas, D., Birkbeck, N., Schmidt, M., Jagersand, M., Murtha, A.: 3D variational brain tumor segmentation using a high dimensional feature set. In: Proc. ICCV, pp. 1–8 (2007)
12. Lefohn, A., Cates, J., Whitaker, R.: Interactive, GPU-based level sets for 3D brain tumor segmentation. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 568–572. Springer, Heidelberg (2003)
13. Ho, S., Bullitt, E., Gerig, G.: Level-set evolution with region competition: automatic 3D segmentation of brain tumors. In: Proc. ICPR, pp. 532–535 (2002)
14. Gorlitz, L., Menze, B.H., Weber, M.A., Kelm, B.M., Hamprecht, F.A.: Semi-supervised tumor detection in magnetic resonance spectroscopic images using discriminative random fields. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 224–233. Springer, Heidelberg (2007)
15. Lee, C., Wang, S., Murtha, A., Greiner, R.: Segmenting brain tumors using pseudo-conditional random fields. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 359–366. Springer, Heidelberg (2008)
16. Wels, M., Carneiro, G., Aplas, A., Huber, M., Hornegger, J., Comaniciu, D.: A discriminative model-constrained graph cuts approach to fully automated pediatric brain tumor segmentation in 3D MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 67–75. Springer, Heidelberg (2008)
17. Hunter, D., Lange, K.: A tutorial on MM algorithms. American Statistician 58, 30–37 (2004)
18. Jordan, M.I., Gaharamani, Z., Jaakola, T.S., Saul, L.K.: An introduction to variational methods for graphical models. Machine Learning 37, 183–233 (1999)
19. Dice, L.: Measure of the amount of ecological association between species. Ecology 26, 297–302 (1945)

Spatio-temporal Analysis of Brain MRI Images Using Hidden Markov Models

Ying Wang[1], Susan M. Resnick[2], and Christos Davatzikos[1]

[1] Section of Biomedical Image Analysis, Department of Radiology, University of Pennsylvania, Philadelphia, USA
[2] Laboratory of Personality and Cognition, National Institute on Aging, Baltimore, USA

Abstract. A rapidly increasing number of medical imaging studies is longitudinal, i.e. involves series of repeated examinations of the same individuals. This paper presents a methodology for analysis of such 4D images, with brain aging as the primary application. An adaptive regional clustering method is first adopted to construct a spatial pattern, in which a measure of correlation between morphological measurements and a continuous patient's variable (age in our case) is used to group brain voxels into regions; Secondly, a dynamic probabilistic Hidden Markov Model (HMM) is created to statistically analyze the relationship between spatial brain patterns and hidden states; Thirdly, parametric HMM models under a bagging framework are used to capture the changes occurring with time by decoding the hidden states longitudinally. We apply this method to datasets from elderly individuals, and test the effectiveness of this spatio-temporal model in analyzing the temporal dynamics of spatial aging patterns on an individual basis. Experimental results show this method could facilitate the early detection of pathological brain change.

1 Introduction

A number of advances in medical imaging technologies allow researchers to study the progression of anatomical or functional changes in a number of diseases and therapies. High-dimensionality pattern analysis methods have been increasingly used to measure imaging patterns and use them for individual diagnosis and progression. In the literature of aging and Alzheimer's Disease (AD), high-dimensional classification work aims to provide diagnostic predictors for early marker of AD [1,2,3,4]. However, there is an increasing need for methods that aim to measure subtle gradual progression of change, especially in neurodegenerative diseases. Compared with the dichotomous classification approaches, pattern regression methods offer an alternative approach, which tries to estimate continuous variables from imaging data [5,6,7,8].

Although high-dimensionality pattern analysis methods have potential to provide biomarkers for early detection of Alzheimer's disease (AD), they typically take into account one scan at a time, which can render them less sensitive to subtle longitudinal changes that relate to disease progression. Longitudinal studies

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 160–168, 2010.

allow us to measure subtle changes more accurate by repeatedly evaluating the same subject over time [9,10,11]. For example, Driscoll et al. evaluated 138 non-demented samples with longitudinal scans up to 10 years, and observed that brain volume declined at specific regions for healthy people while accelerated changes were shown in MCI group [11].

To statistically analyze temporal dynamics and capture disease progression, a general solution is spatio-temporal modeling, which has been quite widely used in pattern recognition and computer vision areas, such as speech processing and activity analysis [12,13]. Spatio-temporal analysis has also found applicability in measuring the temporal evolution of brain activation under fMRI studies [14,15,16]. However, spatio-temporal analysis of brain change in longitudinal studies using advanced statistical analysis tools has been relatively scarce. Herein, we propose such an approach based on a popular dynamic model, Hidden Markov Models (HMM). First, we adaptively extract regional features by clustering brain voxels with similar correlation measurements to age. Compared with voxel-wise methods, regional measures can provide more robust and discriminative patterns. Then HMM is constructed to model the temporal evolution of brain change as a sequence of probabilistic transitions from one discrete state to the other. To improve the stability of the methodology, a bagging strategy is adopted to build ensemble HMM models and estimate state path for each subject statistically. Experiments with brain MRI serial scans from older individuals show that the regional feature-based HMM is an effective method to analyze the spatio-temporal change of brain structure. It also can potentially detect abnormal changes due to neurodegeneration, which is accomplished by comparing the individual state trajectory with changes observed in healthy elderly.

2 Material and Methodology

2.1 Materials

9-year longitudinal data with T_1-weighted MRIs from 144 cognitively normal individuals (CN) were used under BLSA study. More image acquisition details about BLSA data are described in [10]. Here, the slope of the California Verbal Learning Test (CVLT) scores for each subject over all years was used to select training samples, because CVLT test has been widely adopted for cognitive performance evaluation. These slopes were calculated by mixed-effects regression, and 58 subjects with the higher and lower CVLT score slopes were chosen for training, while the remaining 86 subjects were used for testing. The characteristics of participants in this study are shown in Table 1. In this work, brain MR scans were pre-processed to three tissue density maps, namely gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF) by extensively validated techniques [2,8]. These three tissue density maps give a quantitative representation of the spatial tissue distribution in a template space. Brain change is reflected by volume values in the respective tissue density maps.

Table 1. Characteristics of participants in the current study

	Training Set	Testing Set
No. of subjects	58	86
Gender: No. of males/females	29/29	49/37
Baseline age (years)	69.63 ± 8.19	71.04±6.82
Age at last visit (years)	76.27 ± 8.31	76.96±7.85
No. of scans		
Year 1	58	86
Year 2	58	86
Year 3	58	76
Year 4	56	72
Year 5	49	68
Year 6	42	65
Year 7	36	59
Year 8	20	53
Year 9	10	39
Total scans	387	620

2.2 Methodology

Regional feature extraction: To generate robust patterns against measurement noise or image pre-processing errors, it is a common approach to group tissue voxels with similar characteristics. As age is the major risk factor for brain change, Pearson correlational analysis of morphological measurements and the corresponding age is adopted to measure the similarity of tissue voxels. However, given the limited number of samples, how to generate the most informative features with good generalizability is still very challenging. Leave-k-out bagging strategy has been proved effective in improving the robustness of measures. Given a training set D of N samples with longitudinal scans $k_n, n = 1, ...N$, bagging procedure generates N Pearson correlation coefficients between tissue values and age, by respectively sampling $D - k_n$ examples from the whole training set with the corresponding replacement. Then **correlation confidence**, defined as the quotient of the mean and variance of these correlation coefficients from the same location u of tissue map i ($i = 1, 2, 3$ represents GM, WM and CSF respectively), is used to evaluate the discrimination ability and robustness of features. The larger the absolute value of this correlation confidence is, the more relevant to brain change this feature is. Here is the mathematical formulation as follows:

$$c^i(u) = \frac{\overline{c_n^i(u)}}{Var(c_n^i(u))} \tag{1}$$

where

$$c_n^i(u) = \frac{\sum_n (f_n^i(u) - \overline{f^i(u)})(y_n - \overline{y})}{\sqrt{\sum_n (f_n^i(u) - \overline{f^i(u)})^2 \sum_n (y_n - \overline{y})^2}} \tag{2}$$

$c_n^i(u)$ is the Pearson correlation coefficient between the tissue values $f_n^i(u)$ and variables y_n^i (age in this work) at location u of tissue map i from the nth leave-k-out case, from which k images of the nth sample are excluded. Here, $\overline{f_i(u)}$ is the mean of $f_n^i(u)$ over all samples, and $\overline{y}$ is the mean of all sample ages y_n. By extensively examining voxels and their respective correlation coefficients, correlation confidence takes into account not only the discriminative ability, but also consistence of feature. The reason is that, outliers can be found via high variance of correlation coefficients, even some correlation coefficients are high at the location u from cross-validation procedure.

We adopted the method of [8] and partitioned the brain into clusters of relatively homogeneous correlation with age. Given these regional clusters, features can be extracted by some statistical analysis on the respective brain regions. In order to produce a small size of effective features for efficient parameter optimization of spatio-temporal model, a feature subset was selected by the ranking criterion, which was the absolute value of leave-k-out correlation confidence defined as formula (1).

Spatio-temporal model: To model the brain structure change over time, it is reasonable to employ a dynamic model with probability distributions, which indicates the spatio-temporal relationship between the observed brain pattern sequences and hidden state trajectories. Since brain change in old adults usually accompanies tissue volume decline irreversibly, Markov process, in which past observations explicitly influence present measurements, is an appropriate representation for the true brain change. Therefore, HMM model with a Markov process combing unobserved state is applied to statistically analyze the spatial brain changes in longitudinal progression, then explain them by the corresponding state at each time point individually.

Mathematically, an HMM is defined by a finite set of J states, and transitions among the states are governed by a set of probabilities called transition probabilities, $\alpha_{i,j} = P(S_t = j|S_{t-1} = i), 1 \leq i, j, \leq J$. Considering the gradual progression of brain structure with age, a continuous probability density function is employed to characterise the relationship between states and brain change measured with regional patterns. Specifically, a weighted sum of M Gaussian distributions is commonly adopted to approximate the probability density function $b_j = \sum_{m=1}^{M} c_{jm} N(\mu_{jm}, \Sigma_{jm}, o_t)$, in which c_{jm} represent weighting coefficients with stochastic constrains $\sum_m^M c_{jm} = 1, 1 \leq j \leq J$, μ and Σ are the mean vectors and covariance matrices. $\pi(j) = P(S_1 = j)$ is the initial probability of hidden states.

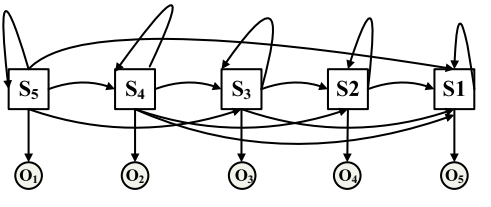


Fig. 1. The left-to-right HMM with 5 states

Our approach aims to characterize progression along a direction, e.g. aging, disease progression or treatment effects. We therefore explicitly incorporate an additional constraint in the model, and introduce a left-to-right HMM structure with 5-states. The state index of left-to-right HMMs decreases or remains the same as time increases, as illustrated in Figure 1. To simplify the model, observation density function b_j is represented by a single Gaussian distribution for each state j in this paper.

A bagging strategy is also used to improve the model generalizability. For each leave-k-out loop, we build a HMM model with the associated parameters (π_n, α_n, b_n). Given the estimated model parameters, Viterbi decoding is employed to find the most likely state path for both training and testing sequences. In order to interpret the state transition, once N state paths are obtained under bagging procedure for each subject, the state that occurs most frequently at each time point is chosen as the final state mode. A framework of the spatio-temporal modelling and analysis is shown in Figure 2.

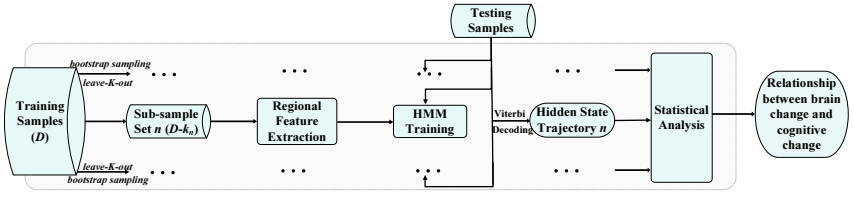


Fig. 2. A bagging framework of HMM modeling and result analysis

3 Results and Discussion

Given the bagging procedure, we summarized the average age distributions of the respective state index values, which were decoded by Viterbi algorithm based on all ensemble HMM models. Overall, the estimated state index value closely correlates with age in both training and testing sets as shown in Figure 3, where the state values decrease with increasing age.

To further evaluate the performance of HMM models, the state transition path for each individual was constructed. We note that almost half of them experienced the state transitions with advancing age, and several representative CNs are illustrated in Figure 4.

In order to investigate the relationship between state transition and cognitive performance decline, CN individuals in the group with positive CVLT score slopes (referred to as Non-Cognitive Decline group, i.e. NCD) were compared with the remaining samples with negative slopes (referred to as Cognitive Decline group, i.e. CD). We examined the number of state transitions for each subject from the training set and the whole set, as illustrated in the first two plots of Figure 5. Though nearly half of them had brain structure change as expected, a much larger proportion of the CN individuals remained structurally stable

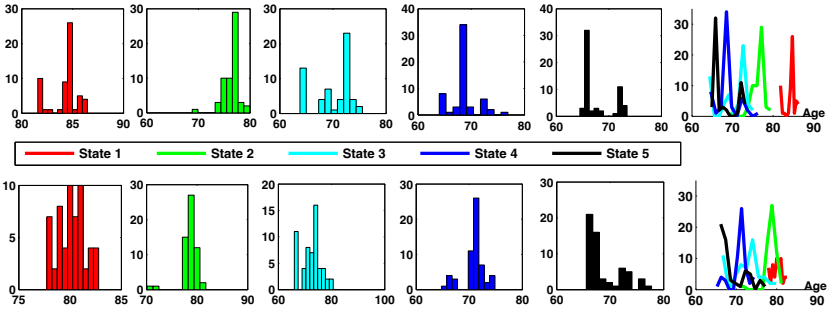


Fig. 3. Average age distributions against the respective hidden states from ensemble HMM models. The top subplots show the average age distributions of training set while the bottom subplots are the results for testing set. The final plots for each set illustrate the combination of individual average age distributions of five states.

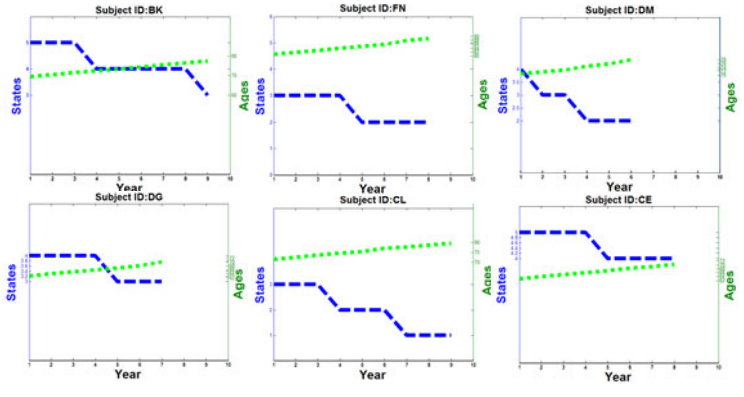


Fig. 4. Individual state paths of several samples in experiments

without any state transition. We also can see that the number of stable subjects in NCD group was larger than that of CD group. To determine which state the stable individuals remained in, we plotted the histogram of the number of state index values for those individuals without transition (the last graph in Figure 5). Notably, what was common in most "CD" subjects was that they showed relatively higher prevalence of states 1-2, which are states with the most abnormal structure, even though they remained relatively stable throughout the follow-up period.

To further understand the difference of the age-related brain changes between "relatively healthy" set and "progressive" set, we examined the average ages for the corresponding states from those two groups, respectively. Here, the samples with relatively minimal CVLT slopes and also high CVLT scores at baseline scan (score > 45) were included into the "relatively healthy" set, while those with rapidly decreasing CVLT slopes were the "progressive" set (slope < -0.5). From

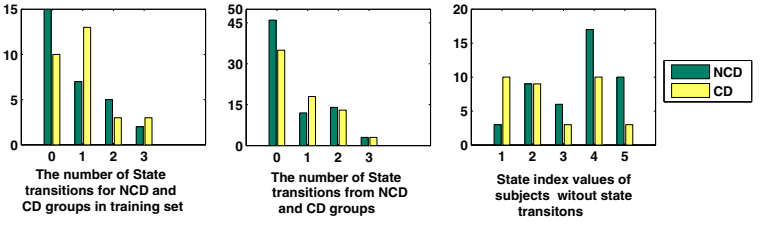


Fig. 5. Statistics of the number of state transitions for NCD and CD groups

Table 2, at states 1 and 2 (which are the most abnormal states), "progressive" group members were younger than those from "relatively healthy" group. It is possible that the "progressive" subjects in these states are the ones with most aggressively evolving disease. In addition, the age difference between the "normal" state 5 and the "abnormal" state 1 in "progressive" group was less than that of the healthy group. For "progressive" group, there was no significant difference of the mean ages between state 2 and 3. This implied that "progressive" individuals progress nonlinearly.

Table 2. Average ages against states

State / Average Age	State 1	State 2	State 3	State 4	State 5
Relatively healthy group	**86.73**	**77.81**	70.54	67.11	65.96
Progressive group	83.34	73.59	71.67	71.16	69.44

Indeed, we found that several individuals demonstrated lower index values of state 1/2 throughout observing period, even though they had high cognitive test scores at baseline, and were relatively younger compared with the average ages of these two states (1/2) as illustrated in Figure 3. Figure 6 shows the state paths with increasing age of two detected samples.

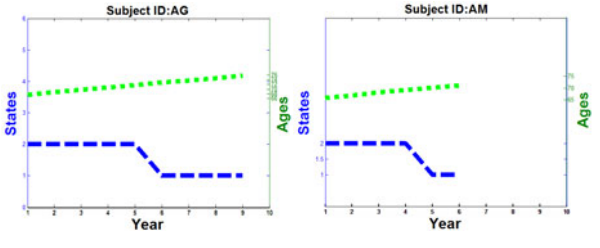


Fig. 6. State paths of two samples with abnormal decline structurally

Therefore, the individuals with early "abnormal" state paths might develop cognitive impairment in the near future. To some extent, this is consistent with

most of MRI studies, i.e., there is an accelerating rate of change among AD-like people, even for those who seem to be cognitively healthy until later disease stages. One of the potentials of HMM models is that they might help identify early state transitions in individuals bound to later develop a disease, and therefore to intervene early enough in the process.

The distinctive patterns of brain change in aging used in HMM modeling are shown in Figure 7. We can see that most of these regional patterns are primarily located at hippocampus, superior temporal gyrus, frontal lobe, cingulate region and precuneus, which are largely in agreement with previous findings [10,11].

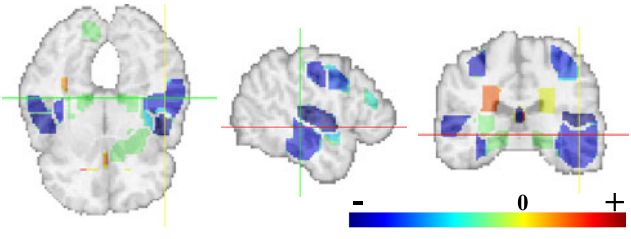


Fig. 7. Regional patterns with brain aging over time are shown by the color-coded ranking score, correlation confidence

4 Conclusion

We presented a methodology for analysis of longitudinal image data, and applied it to serial MRI brain scans of an aging cohort. In view of the statistical analysis between cognitive performance and brain change, we found that considerably more subjects with state transitions came from CD group. Moreover, subjects without state transitions from CD group showed high prevalence of state 1 or 2, which implied their cognitive performance might have already declined before the first visit. In addition, subjects with quickly declining cognitive performance showed faster and nonlinear brain change, compared with "relatively healthy" subjects. These findings are largely consistent with previous studies. Therefore, the HMM model is a promising approach to evaluate the spatio-temporal progression of brain change over time individually. In this work, model parameters were learned automatically from the data. However, the HMM structure was defined manually. This was mainly due to the small size of the data set, which will be addressed in future with the availability of larger data set.

References

1. Golland, P., et al.: Discriminative analysis for image-based studies. In: Dohi, T., Kikinis, R. (eds.) MICCAI 2002. LNCS, vol. 2488, pp. 508–515. Springer, Heidelberg (2002)
2. Fan, Y., et al.: Compare: Classification of morphological patterns using adaptive regional elements. IEEE Transactions on Medical Imaging 26, 93–105 (2007)

3. Kloppel, S., et al.: Automatic classification of MRI scans in Alzheimer's disease. Brain 131(3), 681–689 (2008)
4. Vemuri, P., et al.: MRI and CSF biomarkers in normal, MCI, and AD subjects: Diagnostic discrimination and cognitive correlations. Neurology 73, 287–293 (2009)
5. Duchesne, S., et al.: Relating one-year cognitive change in mild cognitive impairment to baseline MRI features. NeuroImage 47(4), 1363–1370 (2009)
6. Ashburner, J.: A fast diffeomorphic image registration algorithm. NeuroImage 38(1), 95–113 (2007)
7. Franke, K., Ziegler, G., Klöppel, S., Gaser, C.: Estimating the age of healthy subjects from t1-weighted mri scans using kernel methods: Exploring the influence of various parameters. Neuroimage 50(3), 883–892 (2010)
8. Wang, Y., et al.: High-dimensional pattern regression using machine learning: From medical images to continuous clinical variables. NeuroImage 50(4), 1519–1535 (2010)
9. Sullivan, E.V., et al.: Differential rates of regional brain change in callosal and ventricular size: a 4-year longitudinal mri study of elderly men. Cereb. Cortex 12, 438–445 (2002)
10. Resnick, S.M., et al.: Longitudinal magnetic resonance imaging studies of older adults: A shrinking brain. The Journal of Neuroscience 23, 3295–3301 (2003)
11. Driscoll, I., et al.: Longitudinal pattern of regional brain volume change differentiates normal aging from MCI. Neurology 72(22), 1906–1913 (2009)
12. Rabiner, L.R.: A tutorial on hidden markov models and selected applications in speech recognition. In: Proceedings of the IEEE, pp. 257–286 (1989)
13. Moeslund, T.B., et al.: A survey of advances in vision-based human motion capture and analysis. Comput. Vis. Image Underst. 104(2), 90–126 (2006)
14. Katanoda, K., Matsuda, Y., Sugishita, M.: A spatio-temporal regression model for the analysis of functional mri data. NeuroImage 17(3), 1415–1428 (2002)
15. Woolrich, M., et al.: Fully bayesian spatio-temporal modeling of fmri data. IEEE Transactions on Medical Imaging 23(2), 213–231 (2004)
16. Quirós, A., Diez, R.M., Gamerman, D.: Bayesian spatiotemporal model of fmri data. NeuroImage 49(1), 442–456 (2010)

Estimating Local Surface Complexity Maps Using Spherical Harmonic Reconstructions

Rachel Aine Yotter[1], Paul M. Thompson[2], Igor Nenadic[1], and Christian Gaser[1]

[1] Friedrich-Schiller University, Department of Psychiatry,
Jahnstr. 3, 07743 Jena, Germany
[2] Laboratory of Neuro Imaging, UCLA School of Medicine,
Department of Neurology, Los Angeles, CA 90094, USA
{Rachel.Yotter,Igor.Nenadic,Christian.Gaser}@uni-jena.de,
thompson@loni.ucla.edu

Abstract. Cortical surface complexity is a potential structural marker for certain diseases such as schizophrenia. In this study, we developed a measure of fractal dimension (FD) calculated from lowpass-filtered spherical harmonic brain surface reconstructions. A local FD measure was also computed at each vertex in a cortical surface mesh, visualizing local variations in surface complexity over the brain surface. We analyzed the surface complexity for 87 patients with DSM-IV schizophrenia (with stable psychopathology and treated with antipsychotic medication) and 108 matched healthy controls. The global FD for the right hemisphere in the schizophrenic group was significantly lower than that in controls. Local FD maps showed that the lower complexity was mainly due to differences in the prefrontal cortex.

Keywords: surface complexity, fractal dimension, spherical harmonics, schizophrenia, MRI.

1 Introduction

One aspect of brain structure that may be significantly altered in disease is the cortical folding complexity. This can be measured using a metric such as the gyrification index (GI), which is defined as the ratio of the inner surface area (or perimeter in a cross-section) to that of an outer surface convex hull. The GI can be measured in two dimensions by examining cortical slices [1], or in 3D from a reconstructed surface mesh. Differences in cortical folding complexity have been found for some regions in psychiatric disorders such as schizophrenia [2] and bipolar disorder [3]. However, the GI metric has drawbacks, as it depends on the definition of the outer hull, on the plane of section for 2D measures, and may depend on brain size.

These drawbacks can be circumvented using the fractal dimension (FD), which does not rely on the definition of an explicit outer hull (for a review, see [4]). It has been proposed that the brain is a fractal [5], at least over a certain range of scales, and this has been examined using voxel-based information on the overall geometry of the white matter [6, 7]. The FD may also be applied to measure cortical folding complexity. Prior studies found significant differences in FD in psychiatric disorders such as

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 169–176, 2010.

first-episode schizophrenia [8], obsessive-compulsive disorder [9], and Williams syndrome [10], as well as differences associated with sex [11], normal development [12], early-life blindness [13], and IQ [14].

Most definitions of the cortical surface FD rely on the box-counting method, where regional areas are computed at progressively lower sampling resolutions. Since the number of vertices steadily decreases, the position of these vertices can have a large impact on the FD metric and may overlook relevant cortical folding information. This concern may be addressed by aligning sulci across subjects to approximate the same cortical location for each vertex in all subjects [15]. However, alignment is a complicated endeavor that often requires manual delineation of cortical regions.

Surface complexity can also be assessed using spherical harmonic (SPH) expansions. When using SPH reconstructions, the number of vertices (sampling) is the same for all reconstructed surfaces, reducing the influence of individual vertex placement. Furthermore, structural differences in some brain disorders may be easier to detect by investigating the 3D pattern of regional changes rather than a single global metric [16, 17]. For instance, pattern classification techniques can combine signals from different parts of a map to enhance the specificity of morphometric findings [18-22]. This multi-regional information shows promise for developing a more specific brain structural signature of schizophrenia. Using SPH-derived reconstructions, a local FD can be computed at each vertex in the reconstruction, assisting with subsequent pattern-classification approaches.

In this study, we first demonstrate that the fractal dimension values obtained from SPH-derived reconstructions are a valid measure of surface complexity. This is accomplished by measuring the complexity of fractal surfaces with known FD. We then applied complexity analysis to MRI-derived cortical surfaces from schizophrenia patients and healthy controls. We hypothesized that there may be a core pattern of complexity differences that may be detectable in regions commonly implicated in schizophrenia (e.g., hippocampus and dorsolateral prefrontal cortex).

2 Methods

Our computation of the FD for a brain surface mesh has four steps: (1) Generate a surface mesh for each brain hemisphere using the standard FreeSurfer pipeline (http://surfer.nmr.mgh.harvard.edu/). (2) Extract the spherical harmonic coefficients up to a maximum bandwidth (or l-value) of $B = 1536$. (3) Reconstruct 20 brain surface meshes from a progressively increasing series of lowpass-filtered coefficients. (4) Compute global FD using the summed polygon areas of the reconstructed brain surface meshes, and compute local FD using the average area of the neighboring polygons for each vertex. Steps related to complexity measures are detailed below.

2.1 Spherical Harmonic Analysis

To analyze the harmonic content of a surface mesh, the first required step is to re-parameterize the spherical mapping so that it has regularly sampled points with respect to θ and ϕ, where θ is the co-latitude and ϕ is the azimuthal coordinate. The original spherical mapping is from the standard FreeSurfer pipeline. Then, points are generated from equally sampled values of θ and ϕ for all members in the sets, such that there are $2B$ points per set, where B is the bandwidth (or l-value). For each regularly

sampled point, the closest polygon on the spherical mapping is found. Within the closest polygon, a spatial location for the interpolated vertex is approximated using barycentric coordinates. The result is a regularly sampled spherical map in which every point is associated with a coordinate that gives its location on the original surface.

Once the surface mesh is re-parameterized, the harmonic content of a spherical mesh may be obtained using normalized spherical harmonics $Y_l^m(\theta,\phi)$:

$$Y_l^m(\theta,\phi) = P_l^m(\cos\theta)e^{im\phi}, \tag{1}$$

where l and m are integers with $|m| \le l$, and P_l^m is the associated Legendre function defined by:

$$P_l^m(x) = \frac{1}{2^l l!}\left(1-x^2\right)^{\frac{m}{2}} \frac{d^{l+m}}{dx^{l+m}}\left(x^2-1\right)^l. \tag{2}$$

A square-integrable function $f(\theta,\phi)$ on the sphere can be expanded in the spherical harmonic basis such that:

$$f(\theta,\phi) = \sum_{l=0}^{B}\sum_{m=-l}^{l} \left\|Y_l^m\right\|_2^{-2} \hat{f}(l,m)\cdot Y_l^m, \tag{3}$$

where the coefficients $\hat{f}(l,m)$ are defined by $\hat{f}(l,m) = \left\langle f, Y_l^m \right\rangle$ and the L^2-norm of Y_l^m is given by:

$$\left\|Y_l^m\right\|_2^2 = \frac{4\pi}{2l+1}\cdot\frac{(l+m)!}{(l-m)!}. \tag{4}$$

It is possible to solve this system directly by finding the bases first, but a more efficient approach is to use a divide-and-conquer scheme [23].

These coefficients can then be lowpass-filtered, such that only lower coefficients have non-zero values, and passed through an inverse Fourier transform to produce a surface reconstruction. For FD calculations, twenty reconstructions are produced using an upper l-value between 4 and 1536.

2.2 Calculation of Local and Global Complexity Values

Generally, FD is found by finding the slope of a plot regressing log(area) versus log(dimension), over a certain range of scales, where the area is the sum of polygon areas in a given reconstruction. When using spherical harmonic reconstructions, the plot is modified to use bandwidth (or upper l-value), and the slope can be found by regressing log(area) versus log(bandwidth) (Figure 1). Because the area asymptotes for higher-bandwidth reconstructions (e.g., the brain surface is accurately reconstructed if enough coefficients are included), the surface area values included in slope calculations were thresholded at 80% of the original surface area.

In the case of global FD, the area used in the regression is the total area of the reconstructions; for local FD, the area value assigned to a single vertex is the average area of the neighboring polygons. As this area value varies at a local level, the areas

were smoothed using a 25-mm Gaussian heat kernel [24]. Statistical significance was defined using a vertex-level threshold of $p < 0.05$.

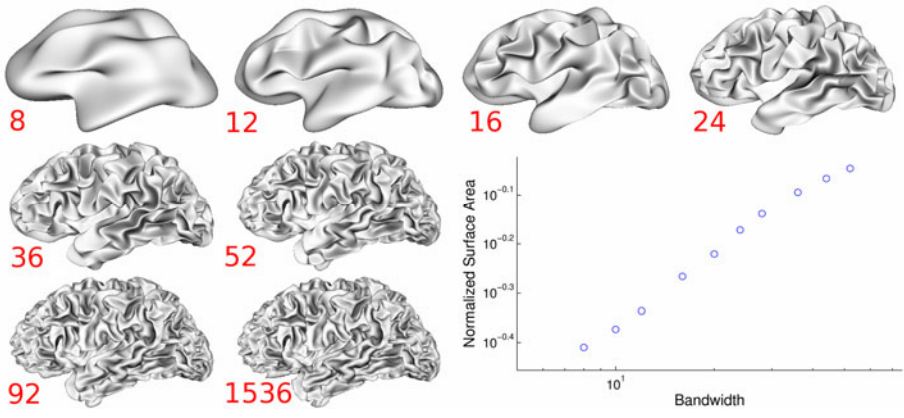


Fig. 1. Fractal dimension is found by finding the slope of a logarithmic plot of surface area versus bandwidth (or upper *l*-value), up to a maximum bandwidth. Surface areas are normalized by the original surface area. A linear approximation is reasonable over this range of scales.

2.3 Artificial Fractal Surfaces

To determine whether the FD values obtained from SPH-derived reconstructions were valid, we generated two sets of von Koch surface meshes that had either a tetrahedral or cubic structure. To avoid self-intersections, the surface meshes slightly deviated from true von Koch surfaces – a reduced length was used for projecting structures. Despite this deviation, however, the two sets of von Koch surfaces still had characteristic FD values determined by measuring the slope of a log-log plot of characteristic dimension versus surface area (Figure 2; cubic: 2.1974, tetrahedral: 2.2936). These surfaces were then processed to generate SPH-derived reconstructions, and the global FD values were extracted from the surface area of the reconstructed surfaces.

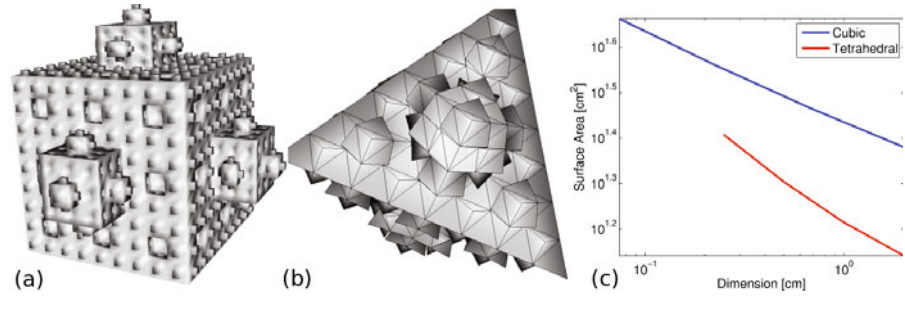


Fig. 2. Modified von Koch fractal surfaces with either a cubic (a) or tetrahedral (b) topology. Each set included surfaces with progressively more detail, obtained by inserting self-similar shapes at lower dimensions. A log-log plot of characteristic dimension versus surface area (c) resulted in a linear line whose slope was the characteristic FD.

2.4 Subject Data

MRI data was acquired from 87 patients (48 male/39 female; mean age = 35.5 years, SD = 11.0) with a DSM-IV diagnosis of schizophrenia and 108 healthy controls (68 male/40 female; mean age = 32.1 years, SD = 10.0). The patients were recruited from the Department of Psychiatry in Jena, Germany, and first screened with a semi-structured interview before being assessed by two psychiatrists establishing the DSM-IV diagnosis. Details of this patient group can be found in [25].

We obtained a high-resolution structural brain MRI from each subject on a 1.5-T Phillips Gyroscan ASCII system using a T1-weighted sequence obtaining 256 sagittal slices covering the entire brain (TR = 13 ms, TE = 5 ms, 25° flip angle, field of view [FOV] = 256 mm, voxel dimensions = 1 × 1 × 1 mm^3) for all subjects. Foam pads were used where appropriate to limit head movement. Prior to image processing, each image was checked manually for artifacts. All scans passed both the manual and automated quality checks.

3 Results

Analysis of the von Koch fractal surfaces using SPH-derived reconstructions resulted in FD values similar to the FD values calculated analytically (cubic: 2.1540 ± 0.007 SEM; tetrahedral: 2.2473 ± 0.007 SEM). Table 1 contains the mean global FD values for the schizophrenic subgroups and controls. Complexity was significantly lower for the right hemisphere in the schizophrenic group.

Table 1. Mean global FD for schizophrenic and control subjects with SEM. *: $p < 0.02$.

	left hemisphere	*right hemisphere*
Control (c)	2.5328 ± 0.0002	2.5340 ± 0.0002
Schizophrenic (s)	2.5355 ± 0.0003	2.5264 ± 0.0002 *

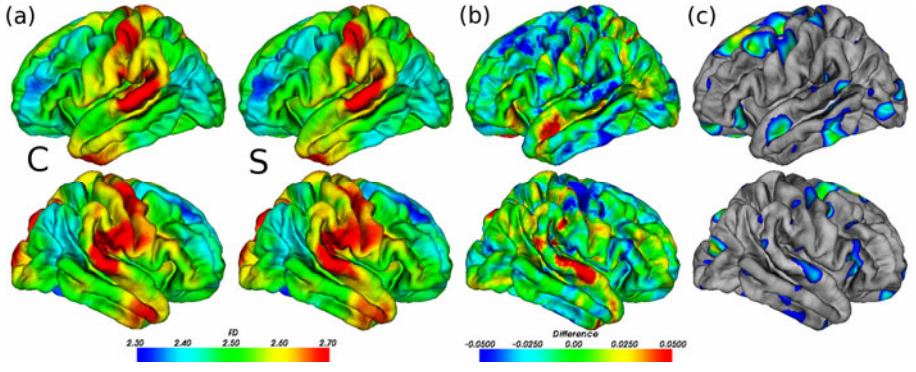


Fig. 3. Local average FD values for control (a, C) and schizophrenic (a, S) groups, for the left (*top row*) and right (*bottom row*) hemispheres. A map of mean complexity differences between groups (b) and *p*-values (c) highlight differences between the groups. In (c), vertices are highlighted if $p < 0.05$, and the values were not corrected for multiple comparisons.

The local FD mapping reveals that the lower overall complexity in the right hemisphere is due to differences in the prefrontal and temporal lobes (Figure 3). Local complexity is also lower for the prefrontal cortex in the left hemisphere.

4 Discussion

We developed a measure of cortical surface complexity that relies on spherical harmonic reconstructions to derive the fractal dimension of brain surfaces both globally and locally (up to single-vertex resolution). Analysis of fractal surfaces demonstrated that this method accurately measures the FD of surfaces. The global FD values for cortical surfaces were similar to previously published FD values [6, 9, 26], indicating that complexity measures of cortical structures based on SPH-derived reconstructions may be an accurate measure of complexity. Other groups have reported much lower FD values. This discrepancy is due to differences in measuring FD (which was usually a box-counting approach applied with relatively low-resolution resampled meshes) or in mathematical definitions of FD measures [8, 10, 11, 15]. The local mapping of complexity was a basic proof-of-concept that lends itself to improvement through inter-subject registration, region-of-interest analysis, and pattern classification methods. Such approaches would allow the extraction of complexity measures for individual lobes and potentially the recognition of schizophrenic patients through structural morphometry alone.

Applied to a large subject pool containing control and schizophrenic subjects, differences in cortical structure were significant in the prefrontal lobe, as predicted. Right hemispheric complexity was lower for the schizophrenic group. There were also significant differences near the hippocampal region, but a more detailed regional analysis needs to be conducted before reaching a conclusion. These findings corroborate earlier hypotheses on aberrant brain development. First, the global FD changes are significant in the right hemisphere, which suggest that the abnormalities are lateralized. Second, the prefrontal and medial temporal localization of FD alterations partially overlaps areas where subtle developmental cellular deficits have been shown [27, 28]. This would suggest that altered FD might be a reflection of abnormal early development of the cortical sheet.

Acknowledgments. This work was supported by the following grants: BMBF 01EV0709 and BMBF 01GW0740.

References

1. Zilles, K., Armstrong, E., Schleicher, A., Kretschmann, H.-J.: The human pattern of gyrification in the cerebral cortex. Anatomy and Embryology 179, 173–179 (1988)
2. Wheeler, D.G., Harper, C.G.: Localised reductions in gyrification in the posterior cingulate: Schizophrenia and controls. Progress in Neuro-Psychopharmacology and Biological Psychiatry 31, 319–327 (2007)
3. McIntosh, A.M., Moorhead, T.W.J., McKirdy, J., Hall, J., Sussmann, J.E.D., Stanfield, A.C., Harris, J.M., Johnstone, E.C., Lawrie, S.M.: Prefrontal gyral folding and its cognitive correlates in bipolar disorder and schizophrenia. Acta Psychiatrica Scandinavica 119, 192–198 (2009)

4. Lopes, R., Betrouni, N.: Fractal and multifractal analysis: A review. Medical Image Analysis 13, 634–649 (2009)
5. Kiselev, V.G., Hahn, K.R., Auer, D.P.: Is the brain cortex a fractal? NeuroImage 20, 1765–1774 (2003)
6. Zhang, L., Dean, D., Liu, J.Z., Sahgal, V., Wang, X., Yue, G.H.: Quantifying degeneration of white matter in normal aging using fractal dimension. Neurobiology of Aging 28, 1543–1555 (2007)
7. Bullmore, E., Brammer, M., Harvey, I., Persaud, R., Murray, R., Ron, M.: Fractal analysis of the boundary between white matter and cerebral cortex in magnetic resonance images: a controlled study of schizophrenic and manic-depressive patients. Psychological Medicine 24, 771–781 (1994)
8. Narr, K.L., Bilder, R.M., Kim, S., Thompson, P.M., Szeszko, P., Robinson, D., Luders, E., Toga, A.W.: Abnormal gyral complexity in first-episode schizophrenia. Biological Psychiatry 55, 859–867 (2004)
9. Ha, T.H., Yoon, U., Lee, K.J., Shin, Y.W., Lee, J.-M., Kim, I.Y., Ha, K.S., Kim, S.I., Kwon, J.S.: Fractal dimension of cerebral cortical surface in schizophrenia and obsessive-compulsive disorder. Neuroscience Letters 384, 172–176 (2005)
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. Journal of Neuroscience 25, 4146–4158 (2005)
11. Luders, E., Narr, K.L., Thompson, P.M., Rex, D.E., Jancke, L., Steinmetz, H., Toga, A.W.: Gender differences in cortical complexity. Nature Neuroscience 7, 799–800 (2004)
12. Blanton, R.E., Levitt, J.G., Thompson, P.M., Narr, K.L., Capetillo-Cunliffe, L., Nobel, A., Singerman, J.D., McCracken, J.T., Toga, A.W.: Mapping cortical asymmetry and complexity patterns in normal children. Psychiatry Research: Neuroimaging 107, 29–43 (2001)
13. Zhang, Y., Jiang, J., Lin, L., Shi, F., Zhou, Y., Yu, C., Li, K., Jiang, T.: A surface-based fractal information dimension method for cortical complexity analysis. In: Dohi, T., Sakuma, I., Liao, H. (eds.) MIAR 2008. LNCS, vol. 5128, pp. 133–141. Springer, Heidelberg (2008)
14. Im, K., Lee, J.-M., Yoon, U., Shin, Y.-W., Hong, S.B., Kim, I.Y., Kwon, J.S., Kim, S.I.: Fractal dimension in human cortical surface: Multiple regression analysis with cortical thickness, sulcal depth, and folding area. Human Brain Mapping 27, 994–1003 (2006)
15. Thompson, P.M., Schwartz, C., Lin, R.T., Khan, A.A., Toga, A.W.: Three-dimensional statistical analysis of sulcal variability in the human brain. Journal of Neuroscience 16, 4261–4274 (1996)
16. Luders, E., Thompson, P.M., Narr, K.L., Toga, A.W., Jancke, L., Gaser, C.: A curvature-based approach to estimate local gyrification on the cortical surface. NeuroImage 29, 1224–1230 (2006)
17. Schaer, M., Cuadra, M.B., Tamarit, L., Lazeyras, F., Eliez, S., Thiran, J.P.: A surface-based approach to quantify local cortical gyrification. IEEE Transactions on Medical Imaging 27, 161–170 (2008)
18. Davatzikos, C., Shen, D., Gur, R.C., Wu, X., Liu, D., Fan, Y., Hughett, P., Turetsky, B.I., Gur, R.E.: Whole-brain morphometric study of schizophrenia revealing a spatially complex set of focal abnormalities. Archives of General Psychiatry 62, 1218–1227 (2005)
19. Kawasaki, Y., Suzuki, M., Kherif, F., Takahashi, T., Zhou, S.-Y., Nakamura, K., Matsui, M., Sumiyoshi, T., Seto, H., Kurachi, M.: Multivariate voxel-based morphometry successfully differentiates schizophrenia patients from healthy controls. NeuroImage 34, 235–242 (2007)

20. Soriano-Mas, C., Pujol, J.s., Alonso, P., Cardoner, N., Menchûn, J.M., Harrison, B.J., Deus, J., Vallejo, J., Gaser, C.: Identifying patients with obsessive-compulsive disorder using whole-brain anatomy. NeuroImage 35, 10328–10327 (2007)
21. Yushkevich, P., Dubb, A., Xie, Z., Gur, R., Gur, R., Gee, J.: Regional structural characterization of the brain of schizophrenia patients. Academic Radiology 12, 1250–1261 (2005)
22. Sun, D., van Erp, T.G.M., Thompson, P.M., Bearden, C.E., Daley, M., Kushan, L., Hardt, M.E., Nuechterlein, K.H., Toga, A.W., Cannon, T.D.: Elucidating a magnetic resonance imaging-based neuroanatomic biomarker for psychosis: Classification analysis using probabilistic brain atlas and machine learning algorithms. Biological Psychiatry 66, 1055–1060 (2009)
23. Healy, D.M., Rockmore, D.N., Moore, S.S.B.: FFTs for the 2-Sphere-Improvements and Variations. Dartmouth College (1996)
24. Chung, M.K., Robbins, S.M., Dalton, K.M., Davidson, R.J., Alexander, A.L., Evans, A.C.: Cortical thickness analysis in autism with heat kernel smoothing. NeuroImage 25, 1256–1265 (2005)
25. Nenadic, I., Sauer, H., Gaser, C.: Distinct pattern of brain structural deficits in subsyndromes of schizophrenia delineated by psychopathology. NeuroImage 49, 1153–1160 (2010)
26. Tang, M., Wang, H.N.: Feature analysis of brain MRI images based on fractal dimension. In: EMBS, Shanghai, pp. 3245–3248 (2005)
27. Eastwood, S.L., Harrison, P.J.: Interstitial white matter neuron density in the dorsolateral prefrontal cortex and parahippocampal gyrus in schizophrenia. Schizophrenia Research 79, 181–188 (2005)
28. Akbarian, S., Bunney Jr, W.E., Potkin, S., Wigal, S., Hagman, J., Sandman, C., Jones, E.: Altered distribution of nicotinamide-adenine dinucleotide phosphate-diaphorase cells in frontal lobe of schizophrenics implies disturbances of cortical development. Archives of General Psychiatry 50, 169–177 (1993)

Brain Morphometry by Probabilistic Latent Semantic Analysis

U. Castellani[1], A. Perina[1], V. Murino[1,2], M. Bellani[2], G. Rambaldelli[3], M. Tansella[3], and P. Brambilla[3,4]

[1] VIPS lab, University of Verona, Italy
[2] Istituto Italiano di Tecnologia (IIT), Italy
[3] Department of Medicine and Public Health, University of Verona, Italy
[4] ICBN Center, University of Udine and Verona, Italy

Abstract. The paper proposes a new shape morphometry approach that combines advanced classification techniques with geometric features to identify morphological abnormalities on the brain surface. Our aim is to improve the classification accuracy in distinguishing between normal subjects and schizophrenic patients. The approach is inspired by natural language processing. Local brain surface geometric patterns are quantized to *visual words*, and their co-occurrences are encoded as *visual topic*. To do this, a generative model, the probabilistic Latent Semantic Analysis is learned from quantized shape descriptors (visual words). Finally, we extract from the learned models a generative score, that is used as input of a Support Vector Machine (SVM), defining an hybrid generative/discriminative classification algorithm. An exhaustive experimental section is proposed on a dataset consisting of MRI scans from 64 patients and 60 control subjects. Promising results are reporting by observing accuracies up to 86.13%.

1 Introduction

Computational neuroanatomy using magnetic resonance imaging (MRI) is a fruitful research field that employs image processing techniques to identify geometric characteristics of different brains [1]. The ultimate goal is to automatically identify structural brain abnormalities by estimating the differences between normal subjects and patients affected by a certain disease. Standard approaches are based on the measurement of volume variations [2] which are useful to explain atrophy or dilation due to illness, but on the other hand, they lack in detecting small structural abnormalities at specific locations. To this aim, more advanced shape analysis techniques have been proposed[3,4,5,6]. In [3] geometric properties are captured by computing spherical harmonic descriptors on brain surfaces. Although results are effective, the method requires shapes registration and data resampling. Such pre-processing is avoided in [4,5], where the so called Shape-DNA signature has been introduced by taking the eigenvalues of the Laplace-Beltrami operator as region descriptor for both external surface and volume. Moreover, possible morphological anomalies can be localized by the

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 177–184, 2010.

analysis of the eigenfunctions. Recently a more general shape analysis technique has been proposed, namely the *feature*-based morphometry (FBM) [6]. FBM identifies anatomical structure that can be used as disease biomarkers without a one-to-one spatial correspondence between all subjects. In order to improve the capability in distinguishing between healthy and non-healthy subjects, learning by example techniques [7] are applied (see for example, [8]). Usually, geometric signatures extracted from the MRI data are used as feature vector for classification purpose [9,10,11]. In [9] a support vector machine (SVM) has been employed to classify cortical thickness which has been measured by calculating the Euclidean distance between linked vertices on the inner and outer cortical surfaces. In [10] a new approach has been defined by combining deformation-based morphometry with SVM. In this fashion, multivariate relationships among various anatomical regions have been captured to characterize more effectively the group differences. Finally, in [11], a unified framework is proposed to combine advanced probabilistic registration techniques with SVM. The local spatial warps parameters are also used to identify the discriminative warp that best differentiates the two groups.

In this paper we proposed a new shape morphometry approach. We combine geometric surface properties extraction with advanced classification techniques. Firstly, local geometric features are computed from the brain surface to characterize single surface vertices. Then, being inspired by the research on natural language processing, geometric features are quantized into local shape configurations by defining the set of *visual words*. In order to improve the shape description, a generative model is learned to find local patterns of co-occurrences, by leading to the definition of the so called *visual topics*. To this aim we employ a generative model, the probabilistic Latent Semantic Analysis (pLSA – [12]). Finally, the topic distributions of each subject are fed to a SVM classifier by defining a proper generative kernel. In this fashion the classifier employs a discriminative-generative approach which represents one of the most effective and advanced classification paradigms. In this work, we focus on schizophrenia by analyzing a dataset of 64 patients and 60 controls. A Region-of-Interest (ROI)-based approach is employed by studying the left-Amygdala. The proposed method is able to satisfy at the same time the main advantages of the previous methods: i) registration is not required as in [4,5,6], since extracted geometric features are position and scale invariant, ii) morphometric abnormalities can be localized and visualized as in [3,11], iii) promising accuracy classification performances are observed by exploiting advanced classification techniques, as suggested in [8]. Moreover, the idea of encoding the geometric pattern co-occurrences as *topics* of a generative model is new for shape analysis and for the biomedical domain, for the best of our knowledge.

2 Geometric Feature Extraction

From the set of 2D ROIs of the left-Amygdala shapes the 3D surface is computed as triangle mesh using marching cubes. A minimal smoothing operation is

applied to remove noise and voxelization effect. We encode geometric properties of the surface using the *Shape Index* [13], which is defined as:

$$si = -\frac{2}{\pi} \arctan\left(\frac{k_1 + k_2}{k_1 - k_2}\right) \quad k_1 > k_2,$$

where k_1, k_2 are the principal curvatures of a generic surface point. The Shape Index varies in $[-1, 1]$ and provides a local categorization of the shape into primitive forms such as spherical cap and cup, rut, ridge, trough, or saddle [13]. Shape index is pose and scale invariant [13] and it has already been successfully employed in biomedical domain [14]. The shape index is computed at each vertex of the extracted mesh. Then, all the values are quantized and an histogram of occurrences is computed. Such histogram represent the descriptor of a given subject and it basically encodes the brain local geometry of a subject, disregarding the spatial relationships. Figure 1 shows the 3D surface of the left-Amygdala

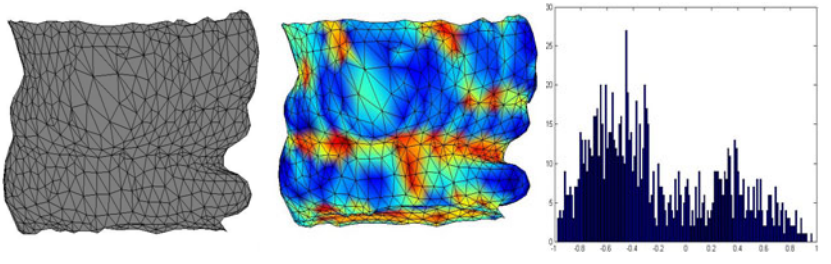


Fig. 1. Geometric feature extraction: 3D surface of the left-Amygdala (left), the surface colored according with Shape Index values (center), and the histogram of Shape Index occurrences (right)

(left), the surface colored according with Shape Index values (center), and the histogram of Shape Index occurrences (right). It is worth noting that convex regions (in blue) are clearly distinguished from concave regions (in red) by the Shape Index values.

3 Topic Models and pLSA

Topic models were introduced in the linguistic scenario, to describe and model documents. The basic idea underlying these methods is that each document is seen as mixture of topics – where a topic is a probability distribution over words. Intuitively, each topic may be related to a particular set of co-occurrent words, that are specific of the treated argument. This representation has one clear advantage: each topic is individually interpretable, providing a probability distribution over words that picks out a coherent cluster of correlated terms. In the shape analysis context, this can be really advantageous since the final goal is to

provide knowledge about morphological abnormalities, and infer possible hidden correlations. In this paper, we focus on probabilistic Latent Semantic Analysis (pLSA)[12] which takes as input a dataset of N documents $\{d_i\}, i$=1,..., N, encoded by set of words. Before applying pLSA, the dataset is summarized by a co-occurrence matrix of size $M \times N$, where the entry $n(w_j, d_i)$ indicates the number of occurrences of the word w_j in the document d_i. The presence of a word w_j in the document d_i is mediated by a latent *topic* variable, $z \in T = \{z_1,..., z_Z\}$, also called *aspect* class, *i.e.*,

$$P(w_j, d_i) = \sum_{k=1}^{Z} P(w_j|z_k)P(z_k|d_i)P(d_i). \tag{1}$$

In practice, the topic z_k is a probabilistic co-occurrence of words encoded by the distribution $P(w|z_k)$, $w = \{w_1,..., w_M\}$, and each document d_i is compactly $(Z < M)$[1] modeled as a probability distribution over the topics, *i.e.*, $P(z|d_i)$, $z = \{z_1,..., z_Z\}$; $P(d_i)$ accounts for varying number of words. The hidden distributions of the model, $P(w|z)$ and $P(z|d)$, are learnt using Expectation-Maximization (EM), maximizing the model data-likelihood L:

$$L = \prod_{i=1}^{N} \prod_{j=1}^{M} P(w_j, d_i)^{n(w_j, d_i)} \tag{2}$$

The E-step computes the posterior over the topics, $P(z|w, d)$, and the M-step updates the parameters, $P^C(w|z)$ which identifies the model. Once the model has been learnt, the goal of inference is to estimate the topic distribution of a novel document. To do this, one can use the standard learning algorithm keeping fixed the parameters $P(w|z)$.

To perform generative classification with pLSA one has to learn a model per-class and assign a new sample to the category whose model fits the point best, i.e., the model with highest likelihood (see Equation 2). Recently, other approaches successfully used meaningful distributions or other by-products coming from a generative model, as feature for a discriminative classifier. The feature vector resulting after this mapping is called *generative score*. The intuition is that generative models like pLSA are built to understand how samples were generated, and they haven't any notion of discrimination; on the other hand, discriminative classifiers are built to separate the data and they are highly more effective if the data has been previously "explained" by a generative model.

Here turn now to explain how we adopt this ideas to the problem in hand: in our case the words w_j are the quantized geometric features previously presented and the documents d are the subjects. The number of bins (words) with which we quantize the continuous shape index values (M=150) has been heuristically chosen in order to get the best possible performance by classifying directly the feature histograms. As result, the input of pLSA are histograms of quantized shape index values and we want to look for co-occurrence of geometric features.

[1] Both Z and M are constants to be a-priori set.

To classify, we learn a pLSA model for normal controls (i.e., $P^C(w|z)$) and one for schizophrenic patients (i.e., $P^S(w|z)$). Instead of comparing the likelihoods of the two models to classify, we extract a score $\phi(d)$ for each document d according its estimated pLSA posteriors distribution, to use it with a discriminative classifier. In forumlae we have that

$$\phi(d) = \left[P(z^C|d), P(z^S|d)\right] = \left[P(z_1^C|d), \cdots P(z_Z^C|d), P(z_1^S|d), \cdots P(z_Z^S|d)\right] \quad (3)$$

Our intuition is that the co-occurrence of geometric features is different between controls and cases. Since the co-occurrences are captured by the topic distributions $P(z^C|d)$ and $P(z^S|d)$, we are defining a meaningful score for discrimination.

4 SVM Classification

One of the most powerful classifier for object recognition is the Support Vector Machine (SVM). SVM constructs a maximal margin hyperplane in a high dimensional feature space, by mapping the original features through a kernel function. Here, the input of the SVM are the score vectors (Eq. 3). In order to compare data, we employed two kind of kernels: the *Histogram Intersection* (HI) kernel and the χ^2-kernel. Given two brains A and B, described by their quantized shape index distributions d^A and d^B, HI and χ^2 kernel functions are respectively defined as:

$$K_{HI}(A,B) = \sum_{i=1}^{2Z} \min\left[\phi_i(d^A), \phi_i(d^B)\right]; \quad K_{\chi^2}(A,B) = \sum_{i=1}^{2Z} \min\left[\phi_i(d^A), \phi_i(d^B)\right] \quad (4)$$

where $\phi_i(d) = P(z_i|d)$, and 2·Z is the length of the score vector.

In this fashion, according to the generative/discriminative classification paradigm, we use the information coming from the generative process as discriminative features of a discriminative classifier. In practice, we measure similarities in the employed topics during the generation of a sample.

5 Results

The dataset used in this work is composed by 64 patients affected by schizophrenia and 60 healthy control subjects, making our dataset much larger than those commonly used in schizophrenia.

MRI scans were acquired using a 1.5 T Siemens Magnetom Symphony Maestro Class, Syngo MR 2002B by facing several methodological issues dictated by international organization to minimize biases and distortions.

We focused our analysis on the left-Amygdala whose abnormal activity is already investigated in schizophrenia. Regions have been manually traced by experts, according to well established medical guidelines. After the geometric feature extraction and pLSA analysis, two generative models are available: one for normal controls and one for schizophrenic patients.

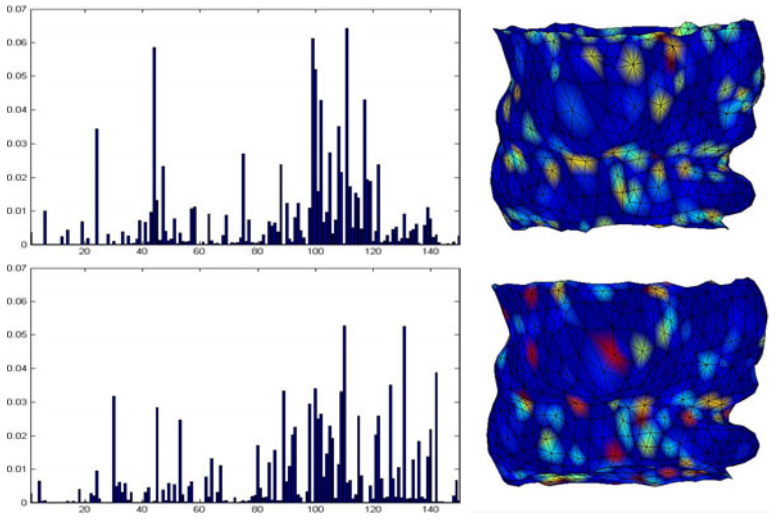


Fig. 2. Output of pLSA: *best*-topic for healthy-model $P^C(w|z = \hat{z_c})$ (top) and patient-model $P^S(w|z = \hat{z_s})$ (bottom). Visual words distribution (left) and their projection onto the surface (right).

In order to obtain a visual feedback of the generative process, we select one high discriminative topic per class, $\hat{z_c}$ and $\hat{z_s}$, respectively for controls and schizophrenic patients. In Figure 2 (left) the word-distribution for both topics is plotted (i.e., the *best*-topic of the healthy model on the top, and of the patient model on the bottom). Note how the visual words are differently generated by the two topics. Moreover, in Figure 2 (right) we show the brain surface of a patient by coloring each vertex according to the relevance of its respective words for the *best*-topic of the healthy model (top) and the patient model (bottom). In this fashion is possible to highlight the surface areas which are more related to the disease, in a likelihood sense (and vice versa for healthy areas).

The classification stage is employed by a random cross-validation strategy. We randomly extracted the 75% of the samples as training set, using the rest for testing. The process is repeated 20 times. Moreover, we investigated the effect of the performance by varying the number of topics Z necessary to capture all the co-occurrences.

Results are reported in Figure 3a. Our best result is 86.13% ±2.17 obtained with 45 topics and histogram intersection kernel. This is impressive if compared with the SVM classification performed directly on the feature histograms $n(w, d)$ by using the same validation strategy and kernel (mean accuracy 58.70% ±9.78) and it shows how beneficial is the proposed generative modeling part.

Although an optimal value of Z should be automatically estimated, our method is robust to the choice of Z. Indeed, we show satisfactory results, i.e., higher than 80% for each $Z \in [25, 55]$.

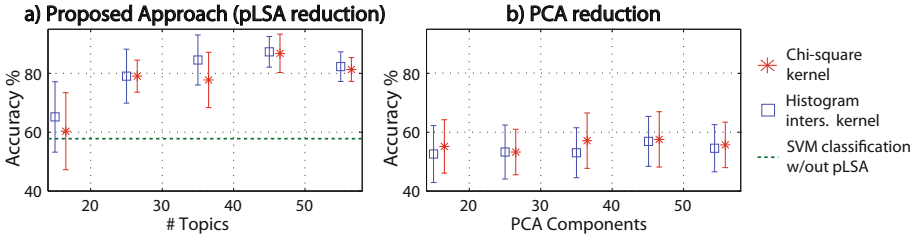


Fig. 3. Error bars showing mean accuracy and standard deviation obtained by varying the number of topics Z. a) The proposed approach. b) PCA dimensionality reduction. Please note how best classification accuracies have lower standard deviations.

As further test, we reduced the dimensionality of the quantized shape index histograms d^A with principal component analysis (PCA), varying the saved components $K \in [25, 55]$. Then, we classified with the same kernels we employed for the proposed approach. Results for PCA, shown in Figure 3b, slightly exceed the chance and they demonstrate how pLSA-based dimensionality reduction is much more discriminative than the one obtained with PCA.

Finally, it is worth noting that our results are in accordance with other similar methods. For instance [11] has shown an accuracy of 90.00% in classifying between 16 first episode schizophrenic and 17 age-matched healthy subjects. In [10], results showed 91.8% of accuracy between 23 cases and 38 controls (only female). Our performance is slightly less but we are analyzing a wider database (no age-sex distinctions).

6 Conclusions

In this paper a new approach for shape analysis is proposed with the aim at characterizing morphological features in schizophrenia. Surface geometric properties are carefully modeled by focusing on left-Amigdala. We have shown that a drastic improvement is observed when the co-occurrences of geometric features are taking into account for classification. In particular, from the pLSA analysis is possible to evaluate the relevance of topics and visual words in modeling both the populations (i.e., healthy and an-healthy people). This open new investigative perspectives in localizing abnormalities due to the presence of multiple geometric factors at the same time. With our method, we have obtained satisfactory results in classification accuracy (i.e., up to 86.13%) which are coherent with previous studies in schizophrenia. Moreover, we have analyzed a larger number subjects in comparison with related work by improving the statistical relevance of the proposed experiments.

Acknowledgements

We acknowledge financial support from the FET programme within the EU-FP7, under the SIMBAD project (contract 213250).

References

1. Giuliani, N.R., Calhouna, V.D., Pearlson, G.D., Francis, A., Buchanan, R.W.: Voxel-based morphometry versus region of interest: a comparison of two methods for analyzing gray matter differences in schizophrenia. Schizophrenia Research 74(2-3), 135–147 (2005)
2. Shenton, M.E., Dickey, C.C., Frumin, M., McCarley, R.W.: A review of mri findings in schizophrenia. Schizophrenia Research 49(1-2), 1–52 (2001)
3. Styner, M., Oguz, I., Xu, S., Brechbuhler, C., Pantazis, D., Levitt, J., Shenton, M., Gerig, G.: Framework for the statistical shape analysis of brain structures using spharm-pdm. In: Open Science Workshop at MICCAI (2006)
4. Niethammer, M., Reuter, M., Wolter, F.E., Bouix, S., Peinecke, N., Koo, M.S., Shenton, M.: Global medical shape analysis using the laplace-beltrami spectrum. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 850–857. Springer, Heidelberg (2007)
5. Reuter, M., Wolter, F.E., Shenton, M., Niethammer, M.: Laplace-Beltrami eigenvalues and topological features on eigenfuntions for statistical shape analysis. Computed-Aided Design 41(10), 739–755 (2009)
6. Toews, M., Wells III, W.M., Collins, D.L., Arbel, T.: Feature-based morphometry. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 109–116. Springer, Heidelberg (2009)
7. Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd edn. John Wiley & Sons, Chichester (2001)
8. Davatzikos, C.: Why voxel-based morphometric analysis should be used with great caution when characterizing group differences. NeuroImage 23(1), 17–20 (2004)
9. Yoon, U., Lee, J., Im, K., Shin, W., Cho, B.H., Kim, I., Kwon, J., Kim, S.: Pattern classification using principal components of cortical thickness and its discriminative pattern in schizophrenia. NeuroImage 34, 1405–1415 (2007)
10. Fan, Y., Shen, D., Gur, R.C., Gur, R.E., Davatzikos, C.: Compare: Classification of morphological patterns using adaptive regional elements. IEEE Transactions on Medical Imaging 26(1), 93–105 (2007)
11. Pohl, K.M., Sabuncu, M.R.: A unified framework for mr based disease classification. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 300–313. Springer, Heidelberg (2009)
12. Hofmann, T.: Unsupervised learning by probabilistic latent semantic analysis. Mach. Learn. 42(1-2), 177–196 (2001)
13. Koenderink, J., van Doorn, A.: Surface shape and curvature scales. Image and Visual Computing 10, 557–565 (1992)
14. Awate, S.P., Yushkevich, P., Song, Z., Licht, D., Gee, J.C.: Multivariate high-dimensional cortical folding analysis, combining complexity and shape, in neonates with congenital heart disease. In: IPMI (2009)

Joint Factor and Kinetic Analysis of Dynamic FDOPA PET Scans of Brain Cancer Patients

N. Dowson[1], P. Bourgeat[1], S. Rose[3], M. Daglish[2,3], J. Smith[2,3], M. Fay[2], A. Coulthard[2,3], C. Winter[2,3], D. MacFarlane[2], P. Thomas[2,3], S. Crozier[3], and O. Salvado[1]

[1] The Australian e-Health Research Centre, RBWH, Brisbane, Australia
nicholas.dowson@csiro.au
[2] Royal Brisbane and Women's Hospital, Herston, Brisbane, QLD, 4029, Australia
[3] University of Queensland, RBWH & St Lucia, Brisbane, Australia

Abstract. Kinetic analysis is an essential tool of Positron Emission Tomography image analysis. However it requires a pure tissue time activity curve (TAC) in order to calculate the system parameters. Pure tissue TACs are particularly difficult to obtain in the brain as the low resolution of PET means almost all voxels are a mixture of tissues. Factor analysis explicitly accounts for mixing but is an underdetermined problem that can give arbitrary results. A joint factor and kinetic analysis is proposed whereby factor analysis explicitly accounts for mixing of tissues. Hence, more meaningful parameters are obtained by the kinetic models, which also ensure a less ambiguous solution to the factor analysis. The method was tested using a cylindrical phantom and the ^{18}F-DOPA data of a brain cancer patient.

1 Introduction

Glioblastoma multiforme, the most common brain tumour [7], is an aggressive malignancy with most patients dying of the disease in less than a year [4]. Treatment generally involves maximal, safe surgical resection of the tumour followed by chemotherapy and adjuvant radiotherapy [7]. Despite this, tumours typically recur within months. The frequency of recurrence could potentially be reduced by specifically targeting regions deemed to contain tumour cells that are likely to proliferate. To delineate the regions containing fecund cells the use of dynamic ^{18}F-DOPA Positron Emission Tomography (PET) is being examined as an addition to the Magnetic Resonance (MR) Images, the current gold standard.

The brain is a particularly difficult organ in which to analyse dynamics. Tissues, such as gray matter, are distributed throughout the volume and have a large area of interface with other tissues. This, combined with the low resolution of PET images means that many voxels consist of multiple tissue types. Hence approaches that rely on selecting a volume of interest (VOI) and fitting a predefined kinetic model [3] are challenging. Such an approach has been used for the brain by Schiepers *et al.*, but only to examine the kinetics of the striatum [10].

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 185–192, 2010.

Factor analysis of dynamic images [1] [2] does not assume a particular model and explicitly accounts for mixtures of tissues within voxels. However, even with a positivity constraint, the solution is neither unique nor simple to locate, necessitating the use of *ad-hoc* explorations of the solution space. Sitek *et al.* [12] proposed a cost function to search the solution space in a more controlled fashion and make the solution more unique, and hence more physically realistic. Even so, factor analysis has been relegated to delineating VOIs in images from which mean time activity curves (TAC) are extracted to estimate certain functions or model parameters [11] [10]. Also, factor analysis attempts to explain all data while ignoring the effects of noise, so the solution is frequently driven by outliers.

Forcing factors or mean TACs to fit a given model can result in more accurate VOIs, as found by Saad *et al.* [9]. Saad used K-Means clustering rather than factor analysis for region segmentation. Clustering and kinetic analysis were iterated in a two step process. However, K-means clustering assumes voxels have a single label, *i.e.* they consist of a single tissue, and cluster centres have limited physical meaning, despite fitting a kinetic model, unless priors such as the distribution of TACs and the volume of particular tissue types are known.

This work proposes combining factor analysis and kinetic analysis. Factor analysis explicitly accounts for mixing of tissues, so physically meaningful parameters are obtained by each kinetic model, while kinetic analysis ensures factors give a solution that is unique. This is achieved by minimising the residual between a model of the brain accounting for several tissue types (one compartment model per tissue) and the TAC associated with each factor. In addition, a cost is associated with what percentage of data is explained by the model based on measured image noise. This allows the problem to be naturally formulated within a single cost function, and no weighting factors need to be selected.

The remainder of the paper is organised as follows. The proposed approach is presented in §2. Some experiments are performed the results of which are discussed in §3 before the paper concludes in §4.

2 Methods

2.1 Factor Analysis

Factor analysis is a statistical method used to describe variability among observations using a small number of unseen variables called factors. Here it is assumed that each tissue type can be represented by a single factor, $\mathbf{f}_p$, where $\mathbf{f}$ is a vector of intensities at different time points and $p \in \{[1; N_p] \cap \mathbb{Z}\}$ is used to index the factors. The time points, t_j, are indexed by $j \in \{[1; N_j] \cap \mathbb{Z}\}$, with frame lengths Δt_j. The voxels are indexed by $i \in \{[1; N_i] \cap \mathbb{Z}\}$. The TAC of each voxel is also represented as a vector, $\mathbf{a}_i$, representing a single observation of N_j variables. The TACs of each voxel are concatenated into an $N_i \times N_j$ matrix, A. Each voxel is formulated as a mixture of factors, *i.e.* $\mathbf{a}_i = \sum_{p=1}^{N_p} c_{ip}\mathbf{f}_p$ or in matrix form:

$$A = MF^T, \tag{1}$$

where M is a $N_i \times N_p$ matrix of mixture coefficients and F is a $N_j \times N_p$ matrix of concatenated factors. Hence *factor analysis* is the process of finding a valid M and F given A, which is an underdetermined problem.

PET data contains substantial amounts of noise, so voxels whose time integral is below some threshold are discarded. Next, the TACs are normalised by their dot product, which is equivalent to removing the proportion of tissue with zero uptake. The normalised TACs are concatenated to form A.

A principal component analysis of A is performed, where each principal component is the eigenvector, ω_p of the covariance matrix of A. Only the N_p-1 eigenvectors with the highest eigenvalues, λ_p, are kept. The mean TAC, $\overline{\mathbf{a}} = \frac{1}{N_i}\sum_i \mathbf{a}_i$, is also retained and concatenated with the the eigenvectors to form W. N_p may be selected according to how many tissues are expected. The eigenvectors and mean TAC cannot be used as factors as they contain negative elements, which have no physical meaning. Hence, an affine transformation of the eigenvectors is used to obtain positive only factors:

$$F = (\mathbf{1}\ R')(\overline{\mathbf{a}}\ \omega_1\ \ldots\ \omega_{N_p-1})^T = RW^T, \tag{2}$$

where R is an $N_p \times N_p$ affine rotation matrix. The values in the R' portion of R encode the vertices of a simplex lying on the hyperplane defined by W. Similarly M becomes a rotation of principal component coefficients: $M = (\mathbf{1}\ L') = LR^{-1}$, where L' is (initially) the $N_i \times (N_p - 1)$ matrix of principal coefficients, so

$$A = LR^{-1}RW^T. \tag{3}$$

Both L and R are optimised so there are $(N_i + N_p)(N_p - 1)$ free parameters to obtain from N_iN_j equations. However A generally only has a rank of around N_p or slightly more if the number of factors was underestimated, the system of equations remains under-determined so any solution will not be unique. Enforcing positivity decreases the size of the solution space. Finding the solution using [1] requires an iterative two step approach. L is updated to ensure positivity and R updated. Next R is updated make F's coefficients positive. This process is dominated by a QR-decomposition each iteration to find $L^{-1}M$, which has complexity $O(N_iN_p^2)$ [6].

Sitek's approach [12] finds M and F directly, but without care it can converge to the mean TAC with the other factors differing only by small amounts, giving the "low" mixing valued by the cost function in the second step.

2.2 Kinetic Analysis

A kinetic model is a linear time invariant model that is fitted to one or more tissues or volumes, which may or may not be independent. Each tissue is assumed to consist of one or more compartments, which are assumed to correspond to physical structures. It is assumed the compartments, C_q, broadly correspond to vascular, interstitial and inter-cellular spaces. The models for each tissue are by necessity an over-simplification of reality. More complex models would overfit the

data resulting in unstable and meaningless model parameters. The parameters are symbolised by k.

The model parameters are obtained by using least squares techniques to minimise the difference between a simulated TAC and data derived TAC. Standard ordinary differential equation (ODE) solving techniques are used. The computational cost for solving an ODE is $O(N_q/\epsilon)$ [13], where ϵ is simulation error, ϵ depends on the amplitude time-step ratio. Hence the complexity is $O(N_q \frac{\Delta t_{\min}}{f_{\max}})$. An example equation for a 3 compartment model is:

$$\begin{pmatrix} k_1 & -k_2 - k_3 & k_4 \\ 0 & k_3 & -k_4 \end{pmatrix} \left(C_1(t)\ C_2(t)\ C_3(t) \right)^T = \frac{\mathrm{d}}{\mathrm{d}t} \begin{pmatrix} C_2(t) \\ C_3(t) \end{pmatrix} \tag{4}$$

Alternatively, spectral analysis techniques may also be used, where a TAC is decomposed into a weighted sum of exponential curves. The exponential parameters and weights can be converted into kinetic parameters and the methods have been shown by Gunn *et al.* to be stable [5], but this technique was deemed too computationally intensive for this work.

For standard least squares techniques an estimate of the (blood) input function, C_1, is required. This may be estimated directly from measurements of the radioactivity of the patient's blood at different timepoints, but this is invasive. Instead, the less accurate procedure of manually selecting a volume within known cerebral vasculature was taken. Here, the first 7 frames (or 90s) of the dynamic sequence are integrated. This trades off the confounding effects of noise (fewer frames) and metabolism (more frames). The vasculature is then segmented by manually thresholding the image and eroding the result by 2 voxels. The transverse sinus is used as the arteries are narrow and partial volume effects occur.

2.3 Joint Factor Analysis and Kinetic Analysis

Factor analysis seeks to obtain a set of pure tissue TACs that lie on a hyperplane in the space of possible TACs. However the locus of TACs obtainable from a given compartment model describes a non-linear k-manifold in TAC-space. If neither system noise nor image noise were present and the system was completely described by N_p compartment models (one model per factor) then the N_p factors would indeed lie on a single $(N_p - 1)$-dimensional hyperplane, with samples distributed between factors due to partial voluming effects, resulting in a set of samples enclosed by a simplex. However real data contains substantial amounts of noise and mixing is generally substantial. Hence hyperplanes frequently do not intersect the k-manifold. Unlike standard factor analysis approaches, this work proposes minimising the distance between each factor on the hyperplane and its closest point on the k-manifold:

$$E_{\text{kinetics}}(p, \mathbf{k}_p) = \frac{1}{N_j} \sum_j [\underbrace{b_p(t_j; \mathbf{k}_p, C_1(t_j))}_{\text{simulated TAC}} - \underbrace{(R_p W^T)_j}_{\text{hyperplane projection}}]^2, \tag{5}$$

where b_p is the simulated TAC as a function of t; $\mathbf{k}_p$ and C_1 are parameters. R is the rotation matrix defined in (3). Each row in R is a single point on the

hyperplane corresponding to the projection of the simulated TAC b_p, and defines one factor as described in §2.1. The points are easily obtained by whitening the vector of simulated TAC values at timepoints t_j: $R \leftarrow \left(\mathbf{1} \;\; W^T \cdot (\mathbf{b}_p - \overline{\mathbf{a}})\right)$.

Without further constraints, this approach is dependent on obtaining a good initial choice of parameters and risks the factor converging to a single point, which would describe the data poorly. Hence, an additional constraint is imposed, requiring some fraction of the data to be enclosed by the simplex defined by the N_p factors:

$$E_{\text{data}} = [[\frac{1}{N_i} \sum_i \phi((\mathbf{a}_i - \overline{\mathbf{a}})^T \cdot W', R')] - \alpha_{\text{non-outliers}}]^2, \tag{6}$$

where ϕ acts as a N_p-dimensional indicator function which returns one for points within the simplex defined by R' and zero otherwise. Note the parts of R and W which are not 1, respectively R' and W' are used. $\alpha_{\text{non-outliers}}$ is the expected proportion of non-outliers, which is estimated by blurring the probability distribution function (PDF) of samples by a Gaussian kernel defined according to the estimated image noise and measuring the proportion of points that move outside the largest connected populated region of the PDF. All data points within the simplex consist of physically meaningful (positive and less than one) proportions of each factor. Data points outside the simplex have physically non-meaningful mixtures that are assumed to arise from noise. The final objective function is:

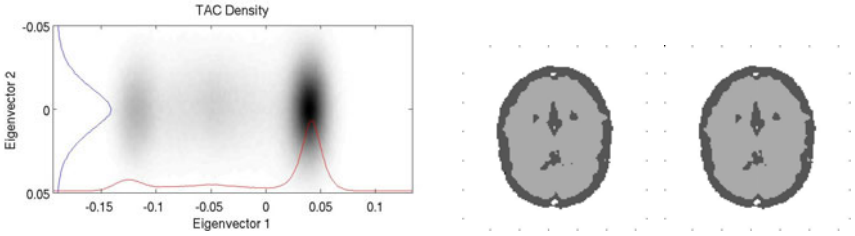
$$\arg\min_{\mathbf{k}} \frac{1}{N_p} \sum_p E_{\text{kinetics}} + E_{\text{data}}. \tag{7}$$

The algorithm is robust to initialisation, if care is taken to ensure R' encompasses most of the data. We select N_p points (factors) most distant from the data mean and from each other. Optimal $\mathbf{k}$-parameters are found for each point using the Nelder-Mead algorithm from 10 random initialisations. The best initialisation is iterated to convergence. Both terms in (7) are between 0 and 1 and have similar scales, hence no weighting parameter is used. The largest computational cost arises from solving one set of ODEs per factor every iteration. Around 2 minutes are needed for a system of 3 factors each comprising 3 compartments.

Unlike standard factor analysis, (7) is driven by selecting kinetic parameters, $\mathbf{k}$, not by selecting R. Noise is explicitly included in the model and it is assumed some data will *not* be completely described by the model, and hence avoids being driven by efforts to ensure noisy outliers consist of mixtures that make physical sense. From the point of view of kinetic analysis, prior regions do not need to be selected, and the data best matching a given kinetic model is found. The final solution is also less ambiguous than the solution obtained by simply ensuring positive mixtures and factors.

3 Experiments and Results

Proving a particular factor analysis method is superior to another is difficult. Ground truth is unavailable for real data, while simulations cannot currently



(a) PDF of TACs (b) Clustering (c) Factor Analysis

Fig. 1. (a) Factor Analysis and (b,c) segmentation of PET SORTEO data

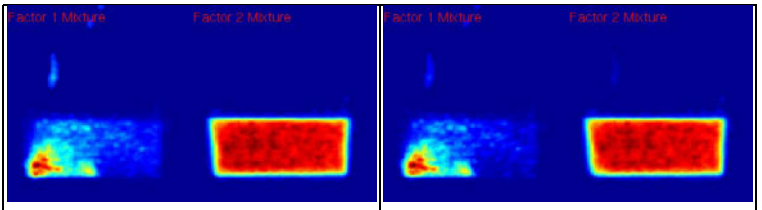
realistically model a true biological system. Even the best available simulations obtained using PET SORTEO, assume only a limited number of factors, differing due to the variation of just one kinetic parameter [8]. This makes testing somewhat artificial as shown in Fig 1. As the TACs in the PET SORTEO data vary across one parameter the locus of the data approximately lies on a line within TAC-space, as shown in Fig. 1a. 3 peaks are observable. Standard clustering should perform better than factor analysis in this simple case. However, the central peak has low density compared to its neighbours so similar segmentations are obtained as shown in Fig. 1b and c.

Hence, two data sets are used. The proposed approach is compared to the method proposed by Di Paola [2]. Future work will also examine the methods proposed by Sitek and Saad.

- A test sequence using a cylindrical phantom containing a liquid solution into which a radioactive tracer was pumped. Although no ground truth was available, the system was simple: only one free parameter exists, which may be extracted in closed form from the TAC within the phantom.
- A brain scan of a patient suffering from cancer, before surgery. The tumour was distant from the striata (which also takes up FDOPA avidly).

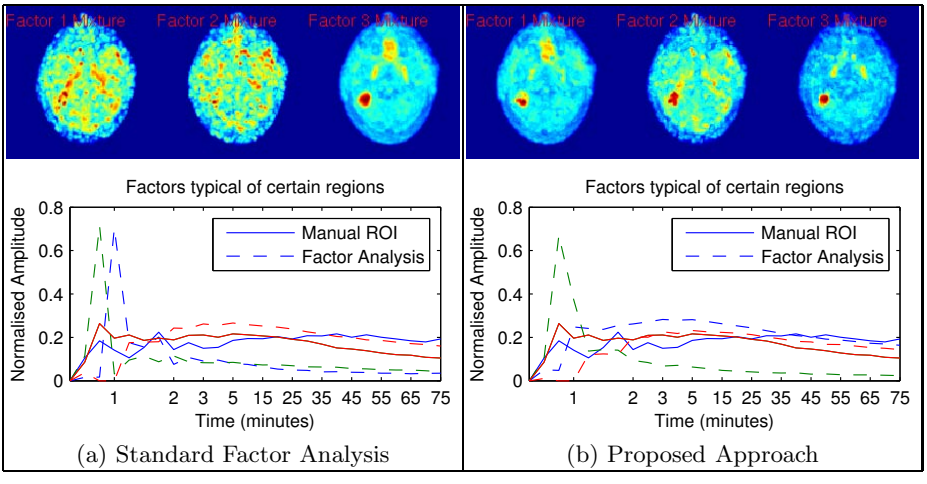
Maximum intensity projections (MIP) of the mixture images from the phantom scans are shown in 2. Images are multiplied by total intensity to prevent noise from obfuscating the result. Both methods identify the input compartment (pipe) and the output compartment (cylinder). Standard factor analysis drives the two factors to the outlying regions of the data. Joint factor and kinetic analysis does not not merely try to encompass all data (and nor does it collapse to a point). Hence it correctly identifies that residual tracer remains in the pipe and has some similarity to the cylinder, giving an above background intensity in the mixture 2 image.

MIPs of the mixtures images for the brain data are shown in Fig. 3, along with the TACs of the 3 factors. Standard factor analysis groups parts of the tumour with the striata, while joint factor and kinetic analysis better separates the two. Both methods separate the tumour into at least two regions, a core and a periphery. Plots on Fig. 3 compare factors to mean TACs of delineations of



(a) Standard Factor Analysis (b) Proposed Approach

Fig. 2. MIP of mixture images, normalised by intensity, for cylindrical phantom



(a) Standard Factor Analysis (b) Proposed Approach

Fig. 3. (Above) MIP of mixture images, normalised by intensity, for pre-surgery test set for 3 factors. (Below) *Three* model derived TACs for Standard factor analysis and the proposed approach (dashed lines) and *two* TACs of the manually delineated cerebellum and striata (solid lines). Striata segmented from static PET image at 50% of maximum. Cerebellum crus obtained by registering PET to 9 labelled atlases and using the majority vote.

the cerebellum crus and the striata. If the joint analysis were better, then two of the factors should better match the two ROIs. In contrast, merely trying to encompass all data gives singular looking curves that do not match ROIs, as shown.

4 Conclusion

This work has proposed a method to jointly perform kinetic analysis and factor analysis of dynamic data. This overcomes the chicken and egg problem of existing methods, namely how to select an unknown region in order to accurately obtain

its characteristic kinetic parameters. A joint approach has the additional advantage of reducing the ambiguity of solutions supplied by existing factor analysis methods. The constraints are not arbitrary, but based on prior knowledge of the system. The proposed approach is efficient because a single cost function is minimised and requires no weight factors to be arbitrarily selected. The proposed method was examined using a real data set and gave hopeful results. Future work will compare the approach to the methods proposed by Sitek and Saad and examine robustness to errors in the estimated arterial input function. The proposed approach will also be extended to constrain the solution further using a Markov Random Field.

References

1. Barber, D.C.: The use of principal components in the quantiative analysis of gamma camera dynamic studies. Phys. Med. Biol. 25(2), 283–292 (1980)
2. di Paola, R., Bazin, J.P., Aubry, F., Aurengo, A., Cavailloles, F., Herry, J.Y., Kahn, E.: Handling of dynamic sequences in nuclear medcine. IEEE. T. Nuclear Science NS-29(4), 1310–1321 (1982)
3. El Fakhri, G., Sitek, A., Guerin, B., Foley Kijewski, M., Di Carli, M.F., Moore, S.C.: Quantitative dynamic cardiac ^{82}rb pet using generalised factor and compartment analyses. J. Nucelar Medicine 46(8), 1264–1271 (2005)
4. Greene, F.L. (ed.): AJCC Cancer Staging Manual, 6th edn., ch. 47, pp. 387–390. Springer, Heidelberg (2002)
5. Gunn, R.N., Gunn, S.R., Turkheimer, F.E., Aston, J.A.D., Cunningham, V.J.: Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling. J. Cerebral Blood Flow & Metabolism 22, 1425–1439 (2002)
6. Knight, P.A.: Fast rectangular matrix multiplication and qr decomposition. Linear Algebra and its Applications 221, 69–81 (1995)
7. Newton, H.B., Jolesz, F.A. (eds.): Handbook of Neuro-Oncology Neuroimaging. Academic Press, London (2007)
8. Reilhac, A., Batan, G., Michel, C., Grova, C., Tohka, J., Collins, D.L., Costes, N., Evans, A.C.: PET-SORTEO: Validation and development of database of simulated PET volumes. IEEE T. Nuclear Science 52(5), 1321–1328 (2005)
9. Saad, A., Hamarneh, G., Moller, T., Smith, B.: Kinetic modeling based probabilistic segmentation for molecular images. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 244–252. Springer, Heidelberg (2008)
10. Schiepers, C., Chen, W., Cloughesy, T., Dahlbom, M., Huang, S.-C.: ^{18}fdopa kinetics and brain tumours. J. Nuclear Medicine 48(10), 1651–1661 (2007)
11. Schiepers, C., Hoh, C.K., Nuyts, J., Wu, H.M., Phelps, M.E., Dahlbom, M.: Factor analysis in prostate cancer: Delineation of organ structures and automatic generation of in- and output functions. IEEE T. Nuclear Science 49(5), 2338–2343 (2002)
12. Sitek, A., Di Bella, E.V.R., Gullberg, G.T., Huesman, R.H.: Correction of ambiguous solutions in factor analysis using a penalized least squares objective. IEEE T. Medical Imaging 21(3), 216–225 (2002)
13. Szczesny, M.: Complexity of initial-value problems for ordinary differential equations of order k. Journal of Complexity 22(4), 514–532 (2006)

Early Detection of Emphysema Progression

Vladlena Gorbunova[1], Sander S.A.M. Jacobs[2], Pechin Lo[1], Asger Dirksen[3], Mads Nielsen[1,4], Alireza Bab-Hadiashar[5], and Marleen de Bruijne[1,6]

[1] Department of Computer Science, University of Copenhagen, Denmark
[2] Department of Biomedical Engineering, Eindhoven University of Technology, The Netherlands
[3] Department of Respiratory Medicine, Gentofte University Hospital, Denmark
[4] Nordic Bioscience A/S, Herlev, Denmark
[5] Swinburne University of Technology, Hawthorn, Australia
[6] Biomedical Imaging Group Rotterdam, Erasmus MC, Rotterdam, The Netherlands

Abstract. Emphysema is one of the most widespread diseases in subjects with smoking history. The gold standard method for estimating the severity of emphysema is a lung function test, such as forced expiratory volume in first second (FEV_1). However, several clinical studies showed that chest CT scans offer more sensitive estimates of emphysema progression. The standard CT densitometric score of emphysema is the relative area of voxels below a threshold (RA). The RA score is a global measurement and reflects the overall emphysema progression.

In this work, we propose a framework for estimation of local emphysema progression from longitudinal chest CT scans. First, images are registered to a common system of coordinates and then local image dissimilarities are computed in corresponding anatomical locations. Finally, the obtained dissimilarity representation is converted into a single emphysema progression score. We applied the proposed algorithm on 27 patients with severe emphysema with CT scans acquired five time points, at baseline, after 3, after 12, after 21 and after 24 or 30 months. The results showed consistent emphysema progression with time and the overall progression score correlates significantly with the increase in RA score.

1 Introduction

Emphysema is one of the most common chronic obstructive pulmonary diseases [1]. It is characterized by irreversible destruction of the lung parenchyma and usually caused by smoking [2].

In clinical practice, the severity of emphysema is commonly assessed using different lung function tests. Along with the lung function tests chest CT scans has been used for diagnosis of emphysema and detection of emphysema progression. The standard CT density scores, such as relative area (RA) below certain threshold, e.g. -950 HU or -930 HU, and the n-th percentile density (nPD) of the lungs, were applied to estimate the emphysema progression [3,4]. CT densitometry scores have shown to be more sensitive measures of emphysema progression than lung function tests [4].

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 193–200, 2010.

One of the major drawbacks of the standard CT density scores is their dependency on the inspiratory level [5,6]. Another important drawback is the lack of sensitivity, since the emphysema progression could only be measured once the intensity of lung tissue decreases below the standard threshold. Texture analysis may resolve this problems. This issue was investigated in a recent study, where a texture-based classification approach was proposed as alternative to the standard emphysema scores [7]. The results showed that the texture-based approach outperforms the RA scores in differentiating diseased from healthy subjects.

Several studies proposed how to estimate disease progression from longitudinal CT scans [5,6]. Authors proposed a method where CT scans are first registered to a common framework and then emphysema progression is estimated based on the average intensity decrease between the two successive scans.

In this paper, we propose a more general way of assessing emphysema progression between a pair of images. First, images are registered to a common system of coordinates. Second, local image histograms at a given location are obtained and dissimilarity measures between the histograms are computed. Thirdly, a measure of progression at the given location is derived from the dissimilarity measures. Finally, an overall disease progression score between the two images is computed. This method is applied to detect emphysema progression in a longitudinal study of patients with Alpha-1 antitrypsin deficiency [4].

2 Method

In this section we describe in details the work flow of the algorithm. The first subsection (2.1) briefly describes the image registration method that was applied to establish the spatial correspondence between images. The following subsection (2.2) presents how local dissimilarities were constructed. The last subsection (2.3) describes how the local disease progression score on subject level was derived from the set of local dissimilarity measures.

2.1 Registration

The image registration framework presented in [6] is used to register the images to a common system of coordinates. The framework starts with a preprocessing step, where the lung fields are extracted from the CT scans and the background value is set to 0 HU. First, an affine transform is applied to correct for global deformations. Then a series of multi-resolution B-Spline transforms with decreasing grid resolution is applied to the affinely registered images. Each transform was optimized using the stochastic gradient descent method.

Finally, the moving image is deformed based on the obtained deformation field. To minimize the intensity differences in the fixed and moving images caused by the difference in respiratory level, the intensities of the deformed image are adjusted with respect to the Jacobian determinant of the deformation field as proposed in [6]. The baseline image I_1 was set as the fixed image, and the four follow up images $I_{2..5}$ were set as the moving images in the registration framework.

2.2 Local Image Dissimilarity

The registration results in dense spatial correspondence, but small misregistrations in the order of 1 mm remain. To minimize the impact of the misregistration, we propose to compare points in the different images using a simplified version of locally orderless images (LOI) [9], where the inner, outer and tonal scales are fixed. A local histogram is constructed using a weighted window function centered around a point x_0. Given an image $I(x_0, \sigma)$ that is observed under the fixed inner scale σ, the LOI at a point x_0 is defined as follows:

$$h_I(i; x_0, \alpha) = \frac{1}{(\sqrt{2\pi}\alpha)^3} \int_0^x A(x, x_0, \alpha) e^{-(I(x,\sigma)-i)^2} dx, \tag{1}$$

where α is the outer scale, which corresponds to the size of the window function $A(x, x_0, \alpha)$ and i is an intensity value. Later in text we denote the histogram using a shorter notation $h = h_I(i; x_0, \alpha)$.

In order to capture different features, in addition to the original image I ($\sigma = 0$), LOIs are also computed from the blurred image and the gradient magnitude. The feature images are all observed under the same scale, which is achieved by blurring the images using a Gaussian kernel with a standard deviation of σ.

Given the two histograms $h_1(i; x)$ and $h_t(i; x)$ obtained in the same anatomical point x from the two images I_1 and I_t respectively, we compute a set of dissimilarity measures $D(I_1, I_t)(x) = \{d_i(h_1, h_t)\}$ between the histograms.

In this paper, we use two classes of dissimilarity measures. The first class consists of L1-norm and Kullback-Leibler divergence between the two histograms $d_1 = ||h_1 - h_t||_{L_1}$, $d_2 = ||h_1 - h_t||_{L_2}$, $d_3 = D_{KL}(h_1, h_t)$. In the second class, the dissimilarity between the local histograms is computed as the difference between the individual measures on each of the histograms $d_i = m_i(h_1) - m_i(h_t)$: the first four moments, the mode, the energy; and the maximum of difference between the cumulative distribution functions of the histograms $d_n = max(cdf(h_1) - cdf(h_t))$ [10].

2.3 Disease Progression Measure

Since LOIs have a certain region of influence, it is not required to compare each and every point in the images. Therefore, a sparse representation of the image is used for comparison instead, where comparison is only performed on a fixed number of regions, N_s, sampled randomly within the image.

For every sample x_i we compute the set of dissimilarity measures $D_I = D(I_1, I_t)$ between the images I_1 and I_t, and the filtered versions of the images $D_G = D(I_{1,\sigma}, I_{t,\sigma})$, $D_{GM} = D(|\nabla I_{1,\sigma}|, |\nabla I_{t,\sigma}|)$. The subscripts I, G, GM denote the original image and response to the Gaussian and Gaussian magnitude filters respectively. Therefore dissimilarity between the two images at the location x_i is defined by the dissimilarity vector $\boldsymbol{D}_{1,t} = \{D_I, D_G, D_{GM}\}_{1,t}$.

The dissimilarity measures from the first class assess the distance between the corresponding local histograms. The dissimilarity measure from the second

class assess the change in the histogram characteristics. If the histograms differ, dissimilarity measures from the first class are strictly positive while the dissimilarity from the second class result in both positive and negative values. We are interested in local changes regardless of the sign therefore only the magnitude of the dissimilarity measures is considered. Finally, the measure of local changes $p_{1,t}(x_i)$ at the sample x_i between the images I_1 to I_t is computed as the L1-norm of the dissimilarity vector, as follow, $p_{1,t}(x_i) = ||\boldsymbol{D}_{1,t}||_{L_1}$.

3 Experiments

3.1 Data

We conducted experiments on subjects with Alpha 1-antitrypsin deficiency monitored during a period of 30 months. A total of 27 subjects were included into the experiments. For each subject low-dose CT images were acquired at five time points: at baseline, after 3, after 12, after 21, and after 24 or 30 months. Out of 27 subjects 11 were scanned after 24 months. The scans were acquired using a tube voltage of 140 kVp, exposure 40 mAs, in-plane resolution 0.78 mm and slice thickness 2 mm without overlap.

Lung function tests were acquired along with the CT scans, of which we used the forced expiratory volume in 1 second (FEV_1). At baseline all the patients performed lung function tests and average FEV_1 for all the subjects was 1.54 ± 0.68 L, and TLC was 8.02 ± 1.57 L, the ratio FEV_1/TLC was 20.27 ± 10.38 %. For the last visit there are 2 missing lung function tests, and the average over the remaining 25 subjects is FEV_1 1.29 ± 0.71L, TLC 7.45 ± 2.51 and ratio FEV_1/TLC 17.93 ± 9.04 %.

3.2 Measuring Local Emphysema Progression

The four follow up CT scans I_2, I_3, I_4, I_5 were first registered to the baseline image I_1. The segmented lung fields from the baseline image I_1 were eroded with a structuring element of size of $3 \times 3 \times 3$ voxels and $N_s = 2000$ positions were randomly sampled from the eroded lung fields. In our experiments we chose the Gaussian scale of the filters $\sigma = 1$ voxel. The radius of the aperture function A was set to $\alpha = 20$ voxels, and the weights were truncated at 3α radius. For the intensity-based histograms the bin width was set to 1 HU resulting in 1000 bins in total in the intensity range from -1000 to 0 HU. For the histograms of the filtered images, the number of bins was set to 1000 and the bin edges were placed uniformly covering the full range of filter responses.

Within a 3 month period changes are expected to be relatively small, therefore the dissimilarities observed in this period reflects mostly image dissimilarity caused by misregistration and interpolation. From this pair of images we obtained the mean and the standard deviation of the dissimilarity vector $\boldsymbol{D}_{1,2}$. Further we normalized all the dissimilarity vectors $\boldsymbol{D}_{1,t=2,3,4,5}$ with respect to the obtained mean and standard deviation and then computed the corresponding progression measures $p_{1,t=2,3,4,5}$.

4 Results

Table 1 reports the summary of the conventional emphysema progression measurements, the decline in FEV_1 (ΔFEV_1 in L) and increase of relative area below the 950HU(ΔRA950 in [%]). The conventional measures were compared with the proposed feature-based disease progression measures. Disease progression measure (PM) on a subject level was computed as the average of dissimilarity measures for all spatial locations. We tested the complete set of dissimilarities (PM (all)) and only Kullback-Leibler divergence between the local histograms of the smoothed images as the local dissimilarity measures (PM (KL)) or only local increase in RA950HU (PM (locΔRA950)). Table 1 presents the average disease progression measures for all consecutive follow up visits. A time trend analysis was performed for the disease progression measurements using a linear mixed model with the time between the baseline and a follow up visit as fixed effect. For the FEV1 we did not conducted time trend analysis because 9 out of 27 subjects had missing FEV1 at least at one of the visits. The t-values are reported in the Table 1. Additionally, correlation coefficients between the progression measured at the last visit by the proposed methods and by conventional emphysema score and lung function are presented in the Table 1.

Table 1. Summary of the disease progression measures. Left part presents the average of the progression measures over all subjects for the follow up visits and the t-value of the time-trend analysis. The correlation coefficients between the progression measures obtained from the last visit with p-values in brackets are presented in the right part of the table.

	Average Progression				Time-trend	Correlation coefficients	
# mnths	3	12	21	30(24)	t-value	ΔRA950	ΔFEV_1
ΔFEV_1	-0.03	0.01	-0.01	-0.13			
ΔRA950	-1.27	0.08	1.33	1.91	6.37		-0.18(0.48)
PM (all)	0.75	0.76	0.89	0.93	3.09	0.51(0.007)	0.11(0.59)
PM (KL)	0.25	0.27	0.30	0.31	7.80	0.45(0.02)	-0.18(0.39)
PM (locΔRA950)	0.0	0.46	0.77	1.03	8.80	0.87(< 0.001)	0.11(0.59)

Fig.1 shows samples locations, indicated with circles, overlaying on the 2D-slices extracted from the registered follow up images. Radius of a circle in the follow up images is proportional to the dissimilarity measure computed from the complete set of dissimilarities. Each row represents different subjects.

In order to investigate local consistency of the local disease progression measures, we tested the simple hypothesis that samples with dissimilarity measure above a threshold T at the previous follow up visit should not decrease the dissimilarity measure in the consecutive visits. The threshold on the dissimilarity measure was selected based on the 25th- and 75th- percentiles, p_{25} and p_{75}, of the dissimilarity measures after the 3 months follow up visit, $T = p_{75} + 1.5(p_{75} - p_{25})$ which corresponds to ~ 2.7 standard deviations. The total number of samples with dissimilarity above the threshold T and the relative percentage of those

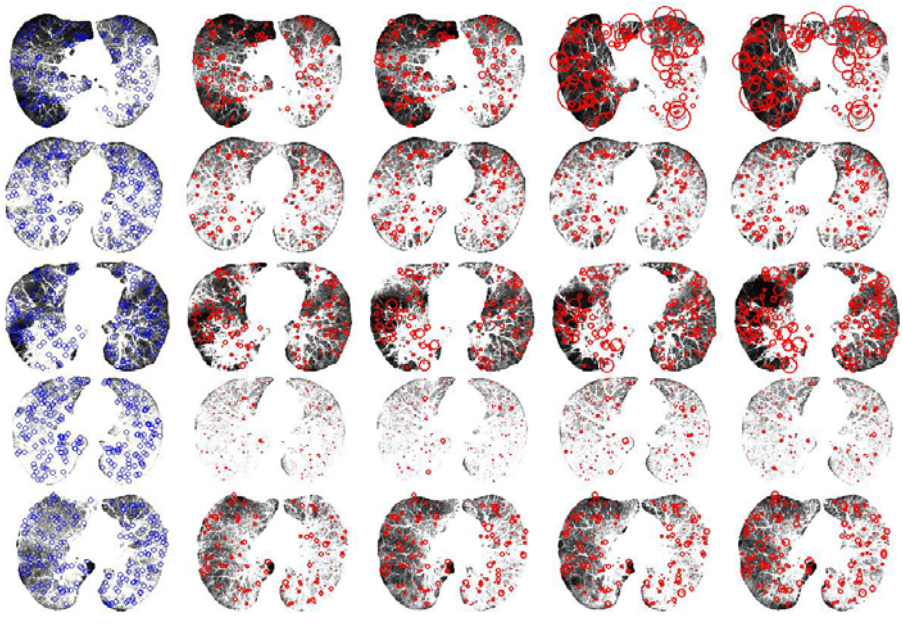


Fig. 1. Rows show mean intensity projection over a stack of 9 sequential slices, selected from different volumetric images. The left most column shows slices extracted from the baseline image, the remaining columns show corresponding slices extracted from the registered 3, 12, 21 and 30(24) months follow up visits, respectively from left to right. All the slices are displayed in the intensity range [-1000,-900]HU. Locations of the random samples (blue and red markers) in the corresponding stack were projected to the image slice. In the follow up images the marker size is proportional to the local dissimilarity measure obtained from the complete set of dissimilarities.

Table 2. Comparison of the local dissimilarity measures. Table presents the overall percentage of samples with dissimilarity measure above the threshold T; in brackets the relative percentage of sampled which increased or preserved the dissimilarity measure above the threshold in all the successive follow up scans.

	Overall percentage [%] (confirmed [%])			
# mnths	3	12	21	30(24)
PM (all)	5.24(41.61)	5.68(61.48)	9.01(75.87)	10.49
PM (KL)	1.19(38.73)	1.75(56.07)	3.72(78.12)	4.93
PM (locΔRA950)	2.86(9.07)	4.38(33.94)	8.88(60.54)	11.48

samples that increase or preserve the same dissimilarity measure in all the successive visits is reported in the Table 2.

Examples of the samples with disease progression measure above the threshold T are presented in the Fig.2. Left plot in Fig.2 shows a subject where most of the samples with the large dissimilarity measure after 3 months were confirmed with all the consecutive follow up scans. Right plot in Fig.2, where the samples did not show consistent dissimilarity measure over time.

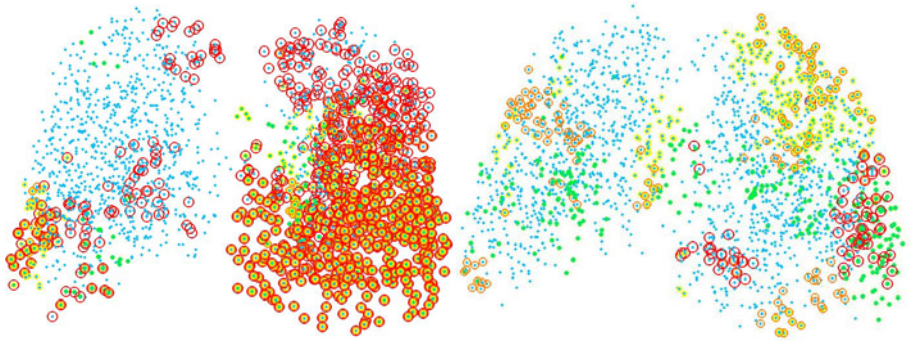


Fig. 2. All selected random samples marked in blue color. Locations with the significantly large dissimilarity measure obtained from the complete set of features at the 3, 12, 21, 30(24) months follow up visits are indicated in green, yellow, orange and red markers respectively.

5 Discussion

In this paper we presented a framework for detection of local emphysema progression. The overall disease progression measure showed significant correlation ($p < 0.01$) with increase in the standard CT score, the relative area below -950HU, between the baseline and the last follow up visit. The correlation with the decline in FEV_1 was not significant for neither the proposed measures nor for the standard CT score. In our dataset the average FEV_1 at baseline was very low, indicating the severity of emphysema already at the baseline visit. This can explain the lack of sensitivity to disease progression of the FEV_1 measurement.

We analyzed time trend based on the conventional emphysema measurements and the proposed dissimilarity-based measurements. The time trend was approximately equally significant for the conventional RA950 disease progression measure, local increase in RA950 and the measure derived from the Kullback-Leibler divergence between local histograms of the smoothed images. The time trend was less significant for the measurement obtained from the complete set of dissimilarities. One of the possible explanations could be sensitivity of the particular dissimilarity measure to the change in image appearance not related to the emphysema progression, for example inflammation or change in local topology like collapsing or appearing bullae. Another possible explanation could be in the construction of the overall combined disease progression score from the complete vector of dissimilarities.

The current drawback of the proposed method is the simplification of the complete dissimilarity vector by its norm. The emphysema is usually characterized by the destruction of the lung tissue thus decreasing image intensities, while inflammation should result in increase of image intensities. In the current framework the two phenomena could result in equal dissimilarity measures. The specific dissimilarity measures such as difference in means of the local histograms is capable of differentiating between the two processes, therefore a

careful investigation of the dissimilarity measures should be done. Furthermore an automatic classification approach could be adapted for this problem, where samples from the image pairs with the 3 months follow up scan represent stable group while samples from the image pairs with the 30(24) months follow up scan represent progressed group.

To conclude, we proposed a method for estimating local disease progression. Results suggested that emphysema progression can be detected before the tissue intensity decreases below the standard CT threshold of -950HU.

Acknowledgments. This work is financially supported by the Danish Council for Strategic Research under the Programme Commission for Nanoscience and Technology, Biotechnology and IT (NABIIT), the Netherlands Organisation for Scientific Research (NWO), and AstraZeneca, Lund, Sweden.

References

1. Rabe, K., et al.: Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease gold executive summary. Am. J. Respir. Crit. Care. Med. 176, 532–555 (2007)
2. Snider, G., Kleinerman, J., Thurlbeck, W., Bengali, Z.: The definition of epmhysema: report of a National Heart, Lung, and Blood Insitute, division of lung diseases workshop. Am. Rev. Resp. Dis. 132, 182–185 (1985)
3. Gietema, H., Schilham, A., van Ginneken, B., van Klaveren, R., Lammers, J., Prokop, M.: Monitoring of smoking-induced emphysema with CT in a lung cancer screening setting: Detection of real increase in extent of emphysema. Radiology 244(3), 890–897 (2007)
4. Dirksen, A., Piitulainen, E., Parr, D., Deng, C., Wencker, M., Stockley, S.S.R.: Exploring the role of CT densitometry: a randomised study of augmented therapy in α_1-antitrypsin deficiency. Eur. Respir. J. 33, 1345–1353 (2009)
5. Staring, M., Bakker, M., Shamonin, D., Stolk, J., Reiber, J., Stoel, B.: Towards local estimation of emphysema progression using image registration. In: Pluim, J., Dawant, B. (eds.) SPIE (Medical Imaging). Proceedings of SPIE, vol. 7259 (February 2009)
6. Gorbunova, V., Lo, P., Ashraf, H., Dirksen, A., Nielsen, M., de Bruijne, M.: Weight preserving image registration for monitoring disease progression in lung CT. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 863–870. Springer, Heidelberg (2008)
7. Sørensen, L., Lo, P., Ashraf, H., Sporring, J., Nielsen, M., de Bruijne, M.: Learning COPD sensitive filters in pulmonary CT. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 699–706. Springer, Heidelberg (2009)
8. Arzhaeva, Y., Prokop, M., Murphy, K., van Rikxoort, E., de Jong, P., Gietema, H., Viergever, M., van Ginneken, B.: Automated estimation of progression of interstitial lung disease in CT images. Medical Physics 37(1), 63–73 (2010)
9. Koenderink, J.J., Van Doorn, A.J.: The structure of locally orderless images. Int. J. Comput. Vision 31(2-3), 159–168 (1999)
10. van Ginneken, B., ter Haar Romeny, B.M.: Multi-scale texture classification from generalized locally orderless images. Pattern Recognition 36(4), 899–911 (2003)

Unsupervised Learning of Brain States from fMRI Data

F. Janoos[1,2], R. Machiraju[1,2], S. Sammet[2], M.V. Knopp[1,2], and I.Á. Mórocz[3]

[1] Dept. of Computer Science, The Ohio State University, USA
[2] Dept. of Radiology, The Ohio State University, USA
[3] Dept. of Radiology, Harvard Medical School, USA

Abstract. The use of multivariate pattern recognition for the analysis of neural representations encoded in fMRI data has become a significant research topic, with wide applications in neuroscience and psychology. A popular approach is to learn a mapping from the data to the observed behavior. However, identifying the instantaneous cognitive state without reference to external conditions is a relatively unexplored problem and could provide important insights into mental processes. In this paper, we present preliminary but promising results from the application of an unsupervised learning technique to identify distinct brain states. The temporal ordering of the states were seen to be synchronized with the experimental conditions, while the spatial distribution of activity in a state conformed with the expected functional recruitment.

1 Introduction

Decoding the cognitive state of a subject from neuroimaging data has always been an important problem in cognitive neuroscience. Recently, this topic has gained prominence with the use of multi-variate pattern recognition (MVPR) to learn complex patterns in fMRI data[1]. Their multivariate nature makes them sensitive to patterns in groups of voxels that individually do not show much structure but as a whole contribute to mentation. Such techniques has been applied to the study of visual[1] and auditory perception, motor tasks[2], word recognition[3], lie-detection, etc., and may have applications in cognitive neuroscience[4], brain-machine interfaces[5], real-time biofeedback[2], etc. The reader is referred to [1] for a plenary review of this topic.

Typically, MVPR in fMRI tries to learn a mapping between the data and labels describing the prevailing attribute of interest (viz. presented stimuli and subject responses) thus only explaining observable attributes of mental processes. However, the identification of covert cognitive states which do not directly relate to experimental conditions is a relatively unexplored problem and could provide important insights into a person's subjective experience from a non-invasive measurement of brain activity. In contrast, sequences of "microstates" have been observed in EEG recordings, ranging from 70ms to 150ms and corresponding to a typical quasi-stable topography of electric field potentials in the brain[6]. They are believed to reflect the activation of different neurocognitive networks, and may be the "atoms of thought" that constitute the seemingly continual "stream of consciousness". While, temporal limitations would prevent access to these faster phenomena, fMRI is suited to reveal relatively longer lasting cognitive phenomena, such as attention, intention, planning, working memory, decision making, etc., that do not necessarily correspond to observable attributes.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 201–208, 2010.

In this paper, we present a method to learn distinct states of activity distribution from fMRI data in an unsupervised fashion. The Earth Mover's Distance (EMD)[7] is used measure the difference between the distributions of activity in the cerebral cortex of two fMRI scans, and computed using a fast approximation method based on recursive aggregation. The EMD is used to derive a diffusion distance, which is aware of the geometry of the underlying low-dimensional manifold spanned by the data. The scans are then grouped using agglomerative hierarchical clustering in this low dimensional space, where each cluster represents a characteristic distribution of activity in the brain. The method was tested on an fMRI study of arithmetical abilities. Clusters corresponding to distinct mental states were identified. The temporal transitions between the states were observed to be synchronized with the experimental conditions, even though no information about the experiment was used. The observed brain activity in each state was found to agree with the expected neural recruitment during the corresponding mental task, thus reaffirming the ability of fMRI to study cognitive states of the brain.

The layout of the rest of the paper is as follows: In Sect. 2, details about the distance metric and clustering are presented, followed by a discussion of the results in Sect. 3. Finally, we conclude with some remarks and outline future directions in Sect. 4.

2 Method

Preprocessing: The acquired data are subjected to routine pre-processing in SPM8[8] (viz. bias-field correction, normalization to an atlas space, segmentation, motion correction and co-registration). The fMRI data are then de-noised using a wavelet-based Wiener filter[9], and further motion correction is performed using spatial ICA[10]. The mean volume of the time-series is then subtracted from each volume, and the white matter is masked out. All further processing is done in the grey matter to reduce the dimensionality of the data and to eliminate confounding influences from the white matter, which theoretically should not exhibit a BOLD effect. Each subject is processed independently.

Earth Mover's Distance: To group the fMRI scans $S(t),\ t = 1...T$ into clusters of characteristic activity signatures, we need a measure of the difference between two scans. The Earth Mover's Distance[7] is well-known metric to compare two signatures P and Q defined over an arbitrary domain $\mathbb{X}$, given the ability to measure distances $d(x, y),\ \forall\ x, y \in \mathbb{X}$. It is defined in terms of an optimal flow $f^* : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{R}$, that minimizes the *transportation problem* $f^* = \arg\min_f \sum_{x,y \in \mathbb{X}} f(x, y)d(x, y)$, subject to the constraints: (i) $f(x, y) \geq 0$; (ii) $\sum_y f(x, y) \leq P(x)$; (iii) $\sum_x f(x, y) \leq Q(y)$; (iv) $\sum_{x,y} f(x, y) = \min\{\sum_x P(x), \sum_y Q(y)\}$, where $x, y \in \mathbb{X}$. The Earth Mover's distance is then $EMD(P, Q) = (\sum_x \sum_y f^*(x, y)d(x, y))/(\sum_x \sum_y f^*(x, y))$. Since the domain $\mathbb{X}$ is the set of voxels in the grey matter, the geodesic distance $d(x, y)$ between two voxels $x, y \in \mathbb{X}$ is used. This is defined as the shortest path length on a graph $\mathbb{G}(\mathbb{X}, \mathbb{E})$, with grey matter voxels as nodes, and $(x, y) \in \mathbb{E}$ if and only if x and y are spatially adjacent to each other. Edge weights are $w(x, y) = ||x - y||_2$, their Euclidean distance in physical space.

One major strength of the EMD over voxel-wise metrics, is that it allows for partial matches, and small differences between the signatures will result in a small EMD[7].

While there exist efficient algorithms for computing the EMD based on the *transportation problem*, it exhibits a worst-case complexity of $\mathcal{O}(|\mathbb{X}|^3 \log |\mathbb{X}|)$. For an fMRI study with voxel size of $3 \times 3 \times 3\text{mm}^3$, the number of grey matter voxels is $\approx 5 \times 10^4$, giving a running time of $\mathcal{O}(10^{14})$. If the number of scans is T, it will require $T(T-1)/4$ number of comparisons, making the pair-wise EMD computation prohibitively expensive. However, since $d(x, y)$ is now a shortest-path distance, not an Euclidean distance, standard approximations to the EMD are not applicable.

EMD Approximation: The problem is made tractable using a recursive approximation as follows: Let S_1 and S_2 be two scans with $|\mathbb{X}|$ voxels in the grey matter. Starting with $j = J > 1$ create a low resolution graph $\mathbb{G}_j$ of $|\mathbb{X}|/2^j$ nodes. Each node represents a set $\mathbb{I}_n^j,\ n = 1 \ldots |\mathbb{X}|/2^j$ of 2^j voxels, grouped based on their spatial proximity. The EMD is computed between S_1 and S_2 on this low-res graph as $EMD_j(S_1, S_2)$. If the edge weights and vertex values of $\mathbb{G}_j$ are calculated appropriately, then $EMD(S_1, S_2) \geq EMD_j(S_1, S_2)$, i.e. it is a lower bound on the EMD on the original mesh. If $EMD_j(S_1, S_2) \geq \tau_j$, where τ_j is a threshold, then we approximate $EMD(S_1, S_2)$ by $EMD_j(S_1, S_2)$, otherwise set $j = j - 1$ and repeat. Continue until $j = 0$, in which case $EMD(S_1, S_2) = EMD_0(S_1, S_2)$.

Next, we explain how to correctly aggregate a set of vertices $\mathbb{I} \subseteq \mathbb{X}$ into a new vertex x', so that the EMD computed on this reduced graph $\mathbb{G}'(\mathbb{X}', \mathbb{E}')$ is less than that on the original graph $\mathbb{G}(\mathbb{X}, \mathbb{E})$. For explicatory purposes, let $\sum_x S_1(x) = \sum_x S_2(x)$, though the reasoning holds for the general case. Let $S = S_2 - S_1$ be the difference between the two distributions. Then, the optimal flow f^* from S_1 to S_2 on $\mathbb{G}$ is the solution to the transportation problem, subject to the constraints $\sum_y f(y, x) - \sum_y f(x, y) = S(x)$ and $f(x, y) \geq 0$, for all $x, y \in \mathbb{X}$. Now, the total cost can be partitioned as $\sum_{x,y \in \mathbb{X}} f^*(x, y)d(x, y) = \sum_{x,y \in \mathbb{I}} f^*(x, y)d(x, y) + \sum_{x,y \in \mathbb{X}-\mathbb{I}} f^*(x, y)d(x, y) + \sum_{x \in \mathbb{I}, y \in \mathbb{X}-\mathbb{I}} f^*(x, y)d(x, y)$.

Firstly, through a conservation of mass argument, we see that the total flow from $x \in \mathbb{I}$ to all $y \in \mathbb{X}-\mathbb{I}$ must be $\sum_{x \in \mathbb{I}, y \in \mathbb{X}-\mathbb{I}} f(x, y) = \sum_{x \in \mathbb{I}} S(x)$. Let $f^\dagger$ be the optimal solution to the transportation problem on the graph $\mathbb{G}'$ where the value $S(x') = \sum_{x \in \mathbb{I}} S(x)$. Also, let $w'(x', y) = \min_{x \in \mathbb{I}} w(x, y)$, and d' be the shortest path lengths on $\mathbb{G}'$. Therefore, the last term $\sum_{x \in \mathbb{I}, y \in \mathbb{X}-\mathbb{I}} f^*(x, y)\ d(x, y) \geq \sum_{y \in \mathbb{X}'} f^\dagger(x', y)d'(x', y)$. Secondly, it can be proved though contradiction that the first term $\sum_{x,y \in \mathbb{I}} f^*(x, y)d(x, y) \geq \sum_{x,y \in \mathbb{I}} f^\circ(x, y)d(x, y)$, where f° is the solution to the transportation problem restricted to the subgraph $\mathbb{I}$, subject to the constraints: (i) $f(x, y) \geq 0$; (ii) $\sum_y f(x, y) \leq S_1(x)$; (iii) $\sum_x f(x, y) \leq S_2(y)$; (iv) $\sum_{x,y} f(x, y) = \min\{\sum_x S_1(x), \sum_y S_2(y)\} - \sum_x S(x)$, with $x, y \in \mathbb{I}$.

This subproblem could again be solved using the above recursive approximation scheme. However, if $|\mathbb{I}|$ is small enough so that the geodesic and ℓ_2 distances between its vertices are approximately equal, then the approximation $EMD(P, Q) \geq ||\bar{P} - \bar{Q}||_2$ can be used, where $\bar{P}$ and $\bar{Q}$ are the centroids of the signatures P and Q respectively [7].

Therefore, through the recursive approximation scheme we get a lower bound as per $EMD(S_1, S_2) \geq EMD_j(S_1, S_2) + \sum_{n=1}^{|\mathbb{X}|/2^j} EMD[\mathbb{I}_n^j](S_1, S_2)$, where $EMD[\mathbb{I}_n^j](S_1, S_2)$ is the EMD restricted to the subgraph defined by $\mathbb{I}_n^j$.

Diffusion Distance: While the EMD provides a good comparison of scans with similar activity patterns (i.e. low EMD), its suitability to quantify the distance between scans with larger differences is more uncertain, apart from the fact that for such comparisons, we have only an approximate EMD. Then, assuming the accuracy of the EMD only in local neighborhoods on the manifold spanned by BOLD images of brain activity, the data can be embedded into a lower dimension Euclidean space using the concept of *diffusion distances*[11], as follows.

An fMRI volume at each time-point $t = 1 \dots T$ is treated as a vertex on a completely connected graph, specified by the $T \times T$ affinity matrix $\mathbf{W}$, where $\mathbf{W}_{t_1,t_2} = \exp\{-EMD(S(t_1), S(t_2))^2 / 2\sigma^2\}$. The user-defined parameter σ defines a notion of proximity between activation patterns. Let $\mathbf{D}_{t,t} = \sum_{t'=1}^{T} \mathbf{W}_{t,t'}$ be the $T \times T$ diagonal degree matrix. Then $\mathbf{M} = \mathbf{D}^{-1}\mathbf{W}$ is a stochastic matrix (i.e. $\sum_{t'=1}^{T} \mathbf{M}_{t,t'} = 1$) defining a random walk on the graph, with $\mathbf{M}_{t_1,t_2}$ as the probability $p(t_2|t_1)$ for a transition from node $S(t_1)$ to $S(t_2)$.

The probability $p(n, t_2|t_1)$ that the random walk starting at $S(t_1)$ will end at $S(t_2)$ in n–steps is given by $\mathbf{M}^n_{t_1,t_2}$. If the generalized eigen-system of $\mathbf{M} = \Psi\Lambda\Phi'$, with Λ the diagonal matrix of eigenvalues $1 = \lambda_1 \geq \dots \geq \lambda_T \geq 0$, Φ the matrix of right eigenvectors $(\phi_1 \dots \phi_T)$ and Ψ that of left eigenvectors $(\psi_1 \dots \psi_T)$, then $p(n, t_2|t_1) = \phi_1(t_2) + \sum_{j=2}^{T} \lambda_j^n \phi_j(t_2)\psi_j(t_1)$. Note that $\phi_1(t_2)$ is the stationary distribution $\lim_{n\to\infty} p(n, t_2|t_1)$ of the random walk, and is independent of starting point $S(t_1)$.

The diffusion distance between two vertices (i.e. fMRI scans) is defined as :

$$D_n^2(S(t_1), S(t_2)) = \sum_{t=1}^{T} |p(n, t|t_1) - p(n, t|t_2)|^2 \phi_1(t) = \sum_{j=1}^{T} \lambda_j^{2n} (\psi_j(t_2) - \psi_j(t_1))^2.$$

The parameter n defines the scale of the diffusion process and controls the sensitivity of the metric on the local geometry, with smaller n making the metric more sensitive to local differences. It can be seen that $D_n^2(S(t_1), S(t_2))$ is an Euclidean distance if the coordinates of $S(t)$ are given by $(\lambda_j^n \psi_j(t) \dots \lambda_T^n \psi_j(t))$.

The diffusion distance, though not equal to the geodesic distance, is related to the Laplace-Beltrami and Fokker-Planck operators on the manifold underlying the graph, and therefore provides a geometrically aware embedding for functions intrinsically defined on it[11]. Moreover, since the *spectral gap* is usually large, with a few eigenvalues close to 1 and most ≈ 0, the diffusion distance can be well-approximated by only the first $\hat{T} \ll T$ eigenvectors, with error of the order of $\mathcal{O}(\lambda^n_{\hat{T}+1})$.

Hierarchical Clustering: After embedding the volumes in this $\hat{T}$–dimensional Euclidean space, they are grouped into K clusters $\{c_1 \dots c_K\}$ using agglomerative hierarchical clustering. These clusters represent distinctive patterns of distribution of the BOLD signal and are probable indicators of the distinctive modes/states of cognition. Next, each cluster c_k, $k = 1 \dots K$ is labeled with an integer value $0 \leq l_k < K$, using dynamic programming to minimize $\sum_{t=1}^{T-1} |l_k(t+1) - l_k(t)|$, where $l_k(t)$ is the label of cluster c_k if $S(t) \in c_k$. This results in a time series of labels where most transitions are between labels close to each other in value.

3 Results

The method was applied on a data-set of 4 participants, who underwent fMRI while judging the incorrectness of multiplication results. In each trial, two single-digit numbers were either displayed visually or sounded out for 2.5s. After an interval of 0.3s a solution was displayed or sounded for 0.8s. Subjects had *up to* 4s to decide, with a button press, if the solution offered was either: (a) *close* (within $\pm 25\%$ of the correct answer), (b) *too small* or (c) *too big*. The next trial started after a rest of 1s, and each trial ranged from 4s to a maximum of 8.6s.

Acquisition was done on a GE 3T LX scanner with a quadrature head coil using a BOLD sensitized 2D-EPI gradient-echo pulse sequence (TE=35ms, FA=90°, TR=2s, voxel size $3.75 \times 3.75 \times 3.75\text{mm}^3$). A typical session lasted ≈ 18 minutes, with 150 trials and 525 scans. The algorithms were implemented using MATLAB® and Star-P® on an 2.6GHz Opteron cluster with 16 processors and 32GB RAM.

The recursive EMD approximation was done with $J = 10$ and thresholds τ_j were selected adaptively, so that only a small percentage (25%) of the comparisons would need to be performed again at the next level. This resulted in a speed up of $10^3\times$, with an average running time of $\approx$ 23 hrs. per subject. Fig. 1 shows the relative approximation error $(EMD_{true} - EMD_{approx})/EMD_{true}$ with respect to the true EMD. It is observed that the relative error scales linearly with respect to the true distance and is acceptably small for our data-sets.

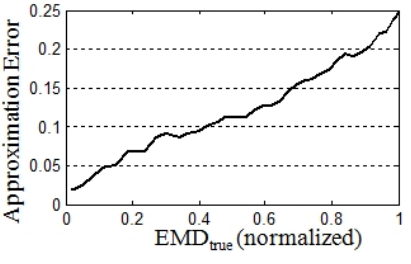


Fig. 1. Relative approximation error $(EMD_{true} - EMD_{approx})/EMD_{true}$ with respect to EMD_{true}. The x–axis is normalized with respect to the maximum EMD.

The parameter σ used in the affinity matrix $\mathbf{W}$ was set such that $\alpha\%$ of all the pair-wise EMDs were less than it, reflecting the assumption that $\sqrt{\alpha}\%$ of scans should be "close" to any given scan. We found the results to be reproducible for $5\% \leq \alpha \leq 20\%$ in our experiments. For the low dimensional embedding in the diffusion metric space, $\hat{T} = 8$ was a conservative value with $\lambda_{\hat{T}+1} < 0.05$. The number of clusters $K = 11$ was selected for all subjects.

In Fig. 2, the median brain-state labels for the audio and visual presentations of the trial are shown, for the four subjects. Here, a strong pattern in the assignment of the state with the phase of the task can be seen. Also, there is a clear separation of labels during presentation of the experiment depending on its modality (audio vs. visual), which converge towards the computation phase, as expected. These findings are especially significant given that no information about the experiment was used when determining the brain state. This synchronization of the brain-state labels with the experiment phase becomes more apparent on examining the intensity distributions for each cluster. The t-score maps for the first subject are shown in Fig. 3. The t-scores at every voxel were computed as within-cluster mean divided by within-cluster standard deviation.

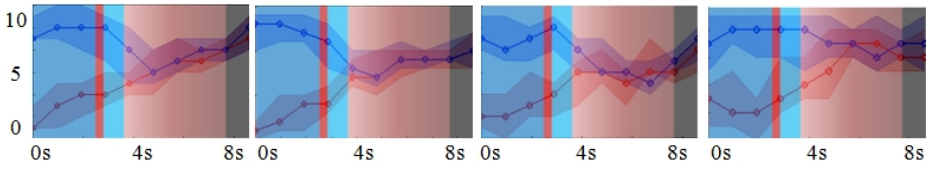


Fig. 2. The median brain-state labels, for all four subjects, during a single trial of the experiment. The phases of the experiment are color-coded to indicate the 2.5s, 0.3s, 0.8s, 0–4s and 1s intervals of each trial. Red and blue lines show the the median brain-states for the visual vs. audio presentation of the numbers, respectively. Also shown are with the 25 and 75 percentile bands.

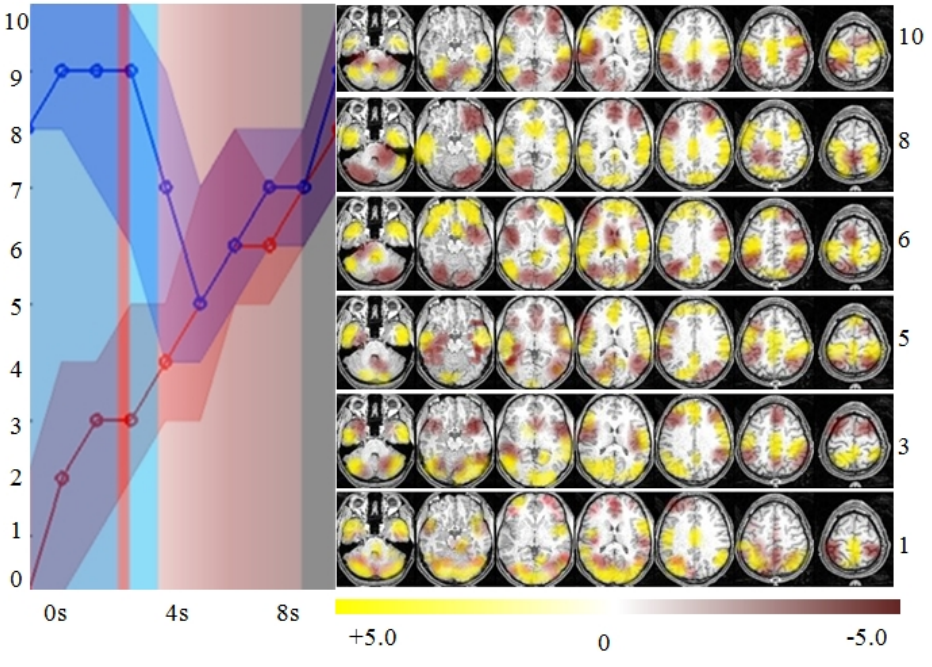


Fig. 3. The within-cluster t–scores for states 1,3,5,6,8,10 for the first subject, overlaid on a high-resolution structural scan. The color coding indicates the intensity in that region. Also shown is the temporal ordering of brain-states for the subject in each trial. The other maps are qualitatively similar and omitted for purposes of concision, while t–scores between [-2,+2] are not displayed for clarity.

State 1 shows strong patterns in the visual cortex, and its occurrence typically corresponds to the visual presentation of the two numbers. State 3, which usually occurs later, is active in visual areas related with size estimation. States 5 and 6, associated with the calculation / judgement phase, are mainly in the frontal and parietal lobes, implicated in higher level cognition and number size assessment. There is also activity in the motor cortex and may be related to the button press. In states 8 and 10, which usually coincide

with the audio presentation of the multiplication problem, the patterns in concentrated in the auditory cortices in the temporal lobe. These findings are in close agreement with those reported[12] for this paradigm using conventional analysis. While there are many unresolved issues in the interpretation of these maps, like negative values and statistical significance, their physiological consistency points to the validity of our method and affirms the potential of such approaches to analyze fMRI data.

4 Conclusion

In this paper, we presented a purely data-driven approach to learn the instantaneous state of the brain from the distribution of intensities in fMRI scans. We used a combination of the Earth Mover's Distance and diffusion distance to induce a metric on the space of fMRI volumes that was aware of its underlying geometry, and performed clustering in this space. We developed a computationally tractable approximation of the EMD based on recursive aggregation. Currently we are working on developing faster and better approximations to the EMD.

The method was applied on a study of arithmetical processing and a pattern in the sequence of brain states was observed that was highly organized with respect to the experiment, with distinct changes from one phase to another. Also the effect of different experimental conditions on the measured response of the brain was observed. Brain maps of activity were obtained that were physiologically meaningful with respect to the expected mental phase, and corresponded with the results of conventional analysis. Though there are many open computational and statistical issues, these results are remarkable given that they were derived solely from the fMRI data, and that no prior information about the experiment or subject behavior was used.

The combination use of EMD and diffusion distance is the main reason the method is able to extract relevant structures from the data. The diffusion distance could be thought of as an operator that uses locally accurate measures of similarity to induce a globally consistent Euclidean metric on the space. For it to succeed however, it is crucial that the underlying measure be accurate when two points are close to each other. In our experiments, we observed that a voxel-wise Euclidean distance (not reported here) performed much worse, as compared to the EMD, resulting in significantly less coherent clusters. This could be because the EMD is less sensitive to small but inevitable differences between signatures corresponding to the same conceptual brain state.

We are currently exploring methods to use these clusters to classify the brain-states of previously unseen scans of the same and of other subjects. Another avenue of research is to learn mappings between the brain states and observed behaviors, allowing for the potential of "brain reading". We are also interested in using information about experimental conditions to improve the identification of the brain-states, using semi-supervised learning algorithms. We believe that such a method coupled with improvements in the temporal resolution of fMRI could give more insight into the temporal organization of the metabolic fingerprints of cognition. Also of interest is the problem of learning the simpler cognitive sub-states, that arise from parallel processes in the different functional circuits of the brain, but which combine in different ways to give rise to the more complex higher level cognition.

References

[1] Haynes, J.D., Rees, G.: Decoding mental states from brain activity in humans. Nat, Rev. Neurosci. 7(7), 523–534 (2006)
[2] LaConte, S.M., Peltier, S.J., Hu, X.P.: Real-time fMRI using brain-state classification. Hum. Brain Mapp. 28(10), 1033–1044 (2007)
[3] Mitchell, T.M., Shinkareva, S.V., Carlson, A., Chang, K.M., 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)
[4] Poldrack, R.A.: The role of fMRI in cognitive neuroscience: where do we stand? Curr. Opin. Neurobiol. 18(2), 223–227 (2008)
[5] Eklund, A., Ohlsson, H., Andersson, M., Rydell, J., Ynnerman, A., Knutsson, H.: Using real-time fMRI to control a dynamical system by brain activity classification. Med. Image Comput. Assist. Interv. 12(Pt 1), 1000–1008 (2009)
[6] Lehmann, D., Pascual-Marqui, R.D., Strik, W.K., Koenig, T.: Core networks for visual-concrete and abstract thought content: a brain electric microstate analysis. Neuroimage 49(1), 1073–1079 (2010)
[7] Rubner, Y., Tomasi, C., Guibas, L.J.: The earth mover's distance as a metric for image retrieval. Int. J. Comput. Vision 40(2), 99–121 (2000)
[8] Friston, K., Holmes, A., Worsley, K., Poline, J., Frith, C., Frackowiak, R.: Statistical parametric maps in functional imaging: A general linear approach. Hum. Brain Map. 2(4), 189–210 (1995)
[9] Alexander, M., Baumgartner, R., Windischberger, C., Moser, E., Somorjai, R.: Wavelet domain de-noising of time-courses in MR image sequences. Magn. Reson. Imaging 18(9), 1129–1134 (2000)
[10] Ng, B., Abugharbieh, R., Mckeown, M.J.: Functional segmentation of fMRI data using adaptive non-negative sparse PCA (ANSPCA), pp. 490–497 (2009)
[11] Coifman, R., Lafon, S.: Diffusion maps. Applied and Computational Harmonic Analysis, Special Issue on Diffusion Maps and Wavelets 21, 5–30 (2006)
[12] Morocz, I., Gross-Tsur, A., von Aster, M., Manor, O., Breznitz, Z., Karni, A., Shalev, R.: Functional magnetic resonance imaging in dyscalculia: preliminary observations. Annals of Neurology 54(S7), S145 (2003)

Geometrical Consistent 3D Tracing of Neuronal Processes in ssTEM Data

Verena Kaynig[1,2], Thomas J. Fuchs[1], and Joachim M. Buhmann[1]

[1] Department of Computer Science, ETH Zurich, Switzerland
verena.kaynig@inf.ethz.ch
[2] Electron Microscopy ETH Zurich (EMEZ), Switzerland

Abstract. In neuroanatomy, automatic geometry extraction of neurons from electron microscopy images is becoming one of the main limiting factors in getting new insights into the functional structure of the brain. We propose a novel framework for tracing neuronal processes over serial sections for 3d reconstructions. The automatic processing pipeline combines the probabilistic output of a random forest classifier with geometrical consistency constraints which take the geometry of whole sections into account. Our experiments demonstrate significant improvement over grouping by Euclidean distance, reducing the split and merge error per object by a factor of two.

1 Introduction

Neuroanatomists build 3d reconstructions of neuronal structures and their synaptic connections in order to gain insight into the functional structure of the brain. As the identification of post synaptic densities is crucial for this task, serial section electron microscopy is the only imaging technique so far, which can provide sufficient resolution. New advances in sample preparation and the imaging process make the acquisition of large data volumes possible [1,2], but the image processing work flow needed to evaluate these data sets still relies heavily on manual labor [3]. This manual intervention renders the process not only error prone and very tedious for the neuroanatomist, but nowadays also becomes a serious bottleneck for the evaluation work flow. In order to build 3d reconstructions of neuronal tissue based on transmission electron microscope (TEM) images, the sample first is embedded into resin, then cut into ultra thin sections of about 50 nm thickness and finally each section is recorded with the microscope. The following image processing work flow consists of aligning the image stack, segmenting structures of interest and building 3d reconstructions out of these segmentations. While the image resolution is dependent on the microscope and can easily achieve 4 nm per pixel, the z-resolution is limited by the section thickness around 40 nm. As a consequence segmentation is usually performed in two dimensions, using the fine resolution to identify membranes of neuronal processes like dendrites and axons [4,5,6,7]. The regions surrounded by the detected membranes then need to be grouped over consecutive sections to extract the geometry of neuronal processes (see Figure 1). Previous work

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 209–216, 2010.

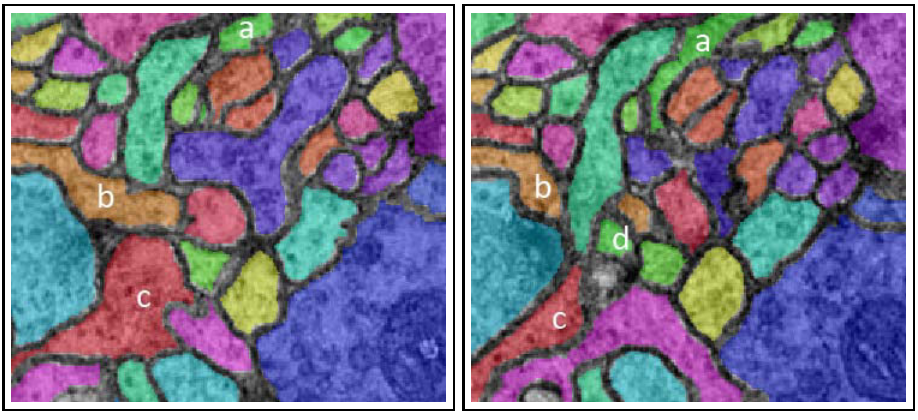


Fig. 1. Example groupings of regions from two adjacent section images (correspondence is indicated by color). The grouping problem is especially hard for thin processes, which have greater flexibility than large structures. In addition structures running longitudinal to the cutting plane, express significant changes in appearance between sections (a-c). Example d has no correspondence in the left section.

has addressed this problem by tracking single processes through the image stack [8,9]. We extend the previous approaches with respect to three important points: (i) instead of tracking single processes the labeling of the whole data volume is optimized, allowing for neuronal processes to start or end inside the volume, (ii) similarity of regions is learned from annotated data, (iii) geometrical consistency between whole sections is taken into account.

2 Method

We regard the problem of three dimensional geometry extraction as partitioning an edge weighted graph into connected components representing an image volume belonging to the same neuronal process. The regions are represented by the vertices V of the graph and the set of edges E connects each region to all regions of the two adjacent sections. Each edge is assigned a weight w_{ij} according to the similarity between regions i and j.

We propose the following processing pipeline to build the edge weight matrix W and to find connected components representing neuronal processes (see Figure 2). First, a set of weight matrices based on features like region overlap or similarity of texture is created. A detailed description of the features is given in Section 2.1. A random forest classifier [10] is trained on manual annotations to predict the similarity between two regions. The weight matrix predicted by the random forest classifier only captures the similarity of pairwise regions. Therefore, a further step refines the weight matrix using geometrical consistent constraints that take the geometry of all neuronal processes included in the section into account. Finally, agglomerative clustering is employed to partition the graph into connected components representing neuronal processes. The hierarchical clustering

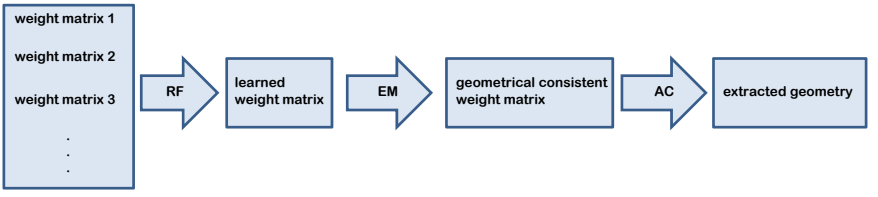


Fig. 2. Processing pipeline for the extraction of 3d geometry of neuronal processes as proposed in this paper. First, a similarity matrix of pairwise regions is learned by a random forest classifier. The learned weight matrix is then combined with geometrical constraints, taking the geometry of all neuronal processes from the whole section into account. Optimization is performed by expectation maximization. Finally agglomerative clustering is used to extract continuous neuronal processes.

scheme starts from individual objects and then progressively merges the regions which are most similar to each other. This system mirrors the approach of the neuroanatomist, who first establishes correspondences between regions that are easy to detect and then refines the partitioning.

2.1 Similarity Features between Regions

The following paragraphs describe the features that are used to train the random forest classifier from manually annotated data. For each feature we build a weight matrix W, each entry representing the edge weight of the corresponding edge in the graph.

Euclidean distance of region center: Each region i is represented by its center of mass $c_i \in \mathbb{R}^3$. The distance of two regions is then given by the Euclidean distance between the two centers:

$$W_{distance}(i,j) = \sqrt{(c_i - c_j) \cdot (c_i - c_j)^T} \tag{1}$$

Overlap of region areas: For each region i, the set P_i contains the position of all pixels belonging to the region ($P_i \in \mathbb{R}^3$). The overlap of two regions is measured by projecting both regions orthogonally to the same plane and building the intersection of both projections:

$$W_{overlap}(i,j) = \#(P_i \cdot A \cap P_j \cdot A), \text{ with } A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{2}$$

Difference in region size: Neuronal processes have only smooth variations in diameter. Therefore the size of corresponding regions should be similar to each other.

$$W_{size}(i,j) = \frac{(\#P_i - \#P_j)^2}{\#P_i + \#P_j} \tag{3}$$

Here $\#P_i$ describes the size of region i in number of pixels. The size difference between two regions is measured by the fraction of the difference in pixels in relation to the total size of both regions. This normalization accounts for the comparability of processes with large or small diameter.

Texture similarity: For the neuroanatomist, texture is an important clue for the extraction of neuronal processes. Intracellular structures like vesicles or microtubules provide information about the type of neuronal process, e.g. bouton or axon, and about the consistent grouping of regions. Following the approach described in [8], we measure the similarity in texture by the cross correlation coefficient of two regions

$$W_{xcorr}(i,j) = X_{\max}(r_i, r_j). \tag{4}$$

Where r_i represents the gray values of region i and $X_{\max}$ denotes the maximal cross correlation between the two regions.

Smooth continuation: This feature weights the connection between two regions i and j according to the smoothest continuation to the next sections. The smoothness of a possible continuation is given by the angle θ_{hij} between the three region centers c_h, c_i and c_j (see Figure 3).

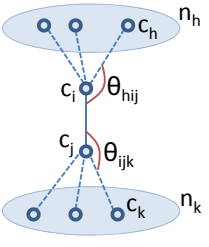


Fig. 3. Illustration of the smooth continuation feature. The smoothness of a possible continuation is given by the function $\theta(c_h, c_i, c_j)$ which measures the angle between the three region centers c_h, c_i and c_j.

$$W_{smooth}(i,j) = \frac{1}{2} \cdot (\min_{h \in n_h} \theta(c_h, c_i, c_j) + \min_{k \in n_k} \theta(c_i, c_j, c_k)), \tag{5}$$

with $\theta(c_h, c_i, c_j) = abs(\pi - \angle(c_h, c_i, c_j))$. The set n_h contains all regions from the section above region i and the set n_k contains all regions from the section below region j. Reflection is employed as border treatment to compute the smooth continuation feature for the first and last section of the stack.

2.2 Geometrical Consistency across Sections

Region correspondences should be assigned in consistency with the overall geometry changes from one section to the next. We address this problem by establishing geometrical consistency of the correspondences between sections. The

approach allows for a non linear but smooth transformation between sections to match correspondent points. Correspondences are not fixed beforehand, but obtained during the optimization [11] For the non-linear transformation we use an explicit polynomial kernel expansion off the points c_i: $\phi(c_i) = [1, c_{i1}, c_{i2}, c_{i1}^2, c_{i1}c_{i2}, c_{i2}^2, \ldots, c_{i2}^d]^T$.

The transformation matrix β projects these points back into the image plane, leading to a nonlinear transformation. Correspondences are assigned by a binary matrix M whose entry m_{ij} is one, if point c_i in one section corresponds to point c_j in the adjacent section and zero otherwise. The energy function to be optimized depends on the similarity of the correspondent regions as classified by the random forest, as well as on the quality of the geometric fit:

$$E(\beta, M) = \sum_{i=1}^{n_i} \sum_{j=1}^{n_j} -m_{ij} ||\phi(c_i)\beta - c_j||^2 + m_{ij} \cdot \ln(W(i,j)) \tag{6}$$

Here the index i runs over the number of regions n_i from one section and j over the number of regions n_j from the adjacent section. The variable m_{ij} contains the associated value of the assignment matrix M and $W(i,j)$ corresponds to the edge weight given by the random forest classifier. Maximizing this energy function can be interpreted as maximizing the data likelihood $p(C_i, C_j|\beta, M)$ where C_i and C_j are matrices containing all points from two adjacent sections. We use expectation maximization to optimize the joint log-posterior, treating the correspondences as unobservable. The algorithm iterates between estimating the expectation of the latent variables m_{ij} while keeping β fix and maximizing the joint log-posterior while keeping the expectation values of M constant.

E-step: In each iteration the variables m_{ij} are replaced by their conditional expectation given β. The expectation values are calculated using the currently optimized β. Under the condition that M is a valid assignment matrix ($\sum_j^{n_2} m_{ij} = 1$, for all $i = 1, \ldots n_2$), we derive the following result:

$$\gamma_{ij} = \mathbf{E}[m_{ij}|C_i, C_j, \beta] = \frac{p(C_i, C_j|\beta, m_{ij} = 1)}{\sum_{l=1}^{n_2} p(C_i, C_j|\beta, m_{il} = 1)} \tag{7}$$

M-step: The expectation of the joint log posterior has the same form as the joint log posterior itself, but with m_{ij} replaced by γ_{ij}. Under the assumption that β is smooth, i.e the components of β are assumed to be normally distributed, maximizing for β yields a weighted ridge regression problem with weights γ_{ij}:

$$\beta \leftarrow (\phi(\widetilde{\mathbf{C_i}})^T \mathbf{\Gamma} \phi(\widetilde{\mathbf{C_i}}) + 2\lambda \mathbf{I})^{-1} \phi(\widetilde{\mathbf{C_i}})^T \mathbf{\Gamma} \mathbf{C_j} \tag{8}$$

where Γ is a $(n_i \cdot n_j) \times (n_i \cdot n_j)$-dimensional diagonal matrix of the weights γ_{ij}. The $(n_i \cdot n_j) \times 2$ matrix $\widetilde{\mathbf{C_i}}$ contains n_j copies of each center point c_i from the first section and the $(n_i \cdot n_j) \times 2$ matrix $\mathbf{C_j}$ contains n_j possible correspondence points from the adjacent section for each point c_i. The parameter λ is the regularization parameter defined by the prior distribution $p(\beta)$. In our experiments λ is set to 0.001.

3 Evaluation

The proposed method is evaluated on ssTEM images, resembling average image quality from neuroanatomy projects. The data set depicts part of the dorsolateral fasciclin-II tract of the ventral nerve cord of the first instar larva of drosophila, at abdominal segment 5. It consists of 30 images with 512x512 pixels. The resolution is 3.7 nm per pixel in the image plane and section thickness is 50nm. The whole data set was annotated exclusively by a neuroanatomist, providing the ground truth for the evaluation. The random forest classifier was trained on this data set using ten fold cross validation to obtain the test error. The remaining pipeline is free of tuning parameters and therefore just applied to the test results of the classifier.

As demonstrated by the plots in Figure 4, each step of our processing pipeline yields significant improvement for the geometry extraction in terms of split and merge error per object. A perfect solution would assign exactly one label per ground truth cluster. For each additional label a split error is counted. A merge error occurs when two ground truth clusters are assigned the same label. If two ground truth clusters are merged more than once, we follow the definition of [12] and count this as one error as the same two objects are involved.

The agglomerative clustering is restricted to establish a maximum of two correspondences for each region, one to the upper and one to the lower section.

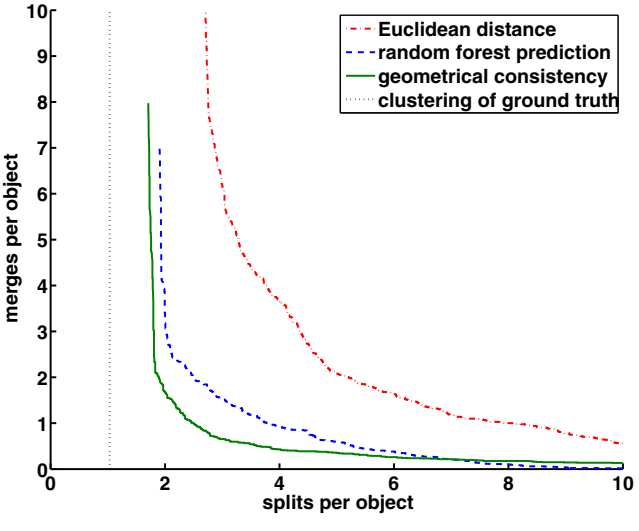


Fig. 4. Evaluation of clustering results according to split/merge error per neuronal process. Depicted are the results for different weight matrices: (i) Euclidean distance of region centers only, (ii) weights learned by the random forest classifier, and geometrical consistent weights. The dotted line corresponds to the best result obtainable without considering branching processes.

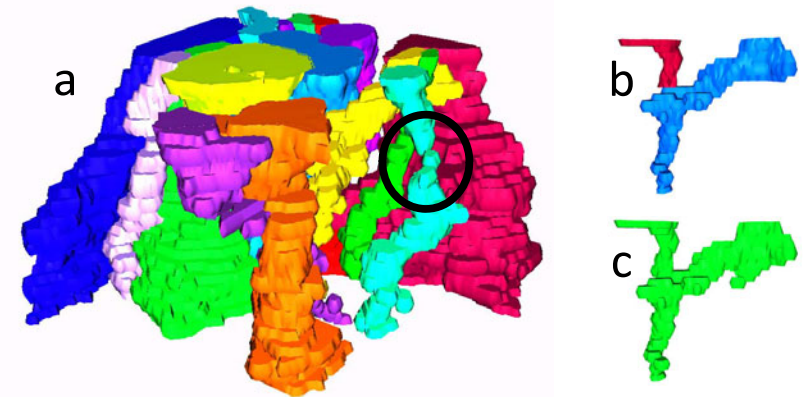


Fig. 5. 3d reconstruction of neuronal processes that were correctly tracked over all 30 sections. A black circle marks an example where regions were correctly grouped despite not having any overlap in adjacent sections. The neuronal process shown in Figure b (ground truth) and c (clustering result) shows an example for a split. The large part including regions moving longitudinal to the cutting direction was correctly grouped and the remaining part was also identified as one object.

Thus, our model allows for starting and ending of new neuronal processes inside the volume, but does not account for branching of processes. The dotted line in Figure 4 marks the best clustering performance achievable by this model.

Examples of extracted geometries are given in Figure 5. The examples demonstrate, that the proposed method is capable of extracting correct geometries also in difficult cases of neuronal processes running longitudinal to the cutting plane and in cases of discontinuities in the geometry due to alignment errors.

4 Conclusion

In this paper we introduced a novel framework for global tracing of neuronal processes in stacks of serial section transmission electron microscopy images. The setting is formulated as a partitioning problem on edge weighted region-graphs. The main contributions of this work are threefold: (i) On the modeling side we propose the use of a random forest classifier to learn a predictor for neighborhood relations of regions within the 3d volume. (ii) Predicted region correspondences are refined taking the geometrical consistency of whole sections into account. (iii) The unsupervised clustering approach results in a non parametric robust procedure for partitioning the graph. In depth evaluation of all single steps of the pipeline and cross validation of the similarity classification demonstrate significant improvement in terms of split and merge error per object. We are convinced that the proposed algorithm is a valuable contribution to the field of neuroscience due to its robustness and general applicability for neuronal process tracing in 3d volumes.

Acknowledgement

We like to thank Nuno Maçarico da Costa, Kevan Martin, and Albert Cardona, Institute of Neuroinformatics UNI-ETH Zurich, for valuable discussions on neuroanatomy and for providing the TEM images.

References

1. Knott, G., Marchman, H., Wall, D., Lich, B.: Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. J. Neurosci. (2008)
2. Denk, W., Horstmann, H.: Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. Curr. Opin. Neurobiol. (2006)
3. Hoffpauir, B., Pope, B., Spirou, G.: Serial sectioning and electron microscopy of large tissue volumes for 3D analysis and reconstruction: a case study of the calyx of held. Nat. Protoc. 2(1), 9–22 (2007)
4. Kaynig, V., Fuchs, T., Buhmann, J.M.: Neuron geometry extraction by perceptual grouping in sstem images. In: CVPR 2010 (2010)
5. Mishchenko, Y.: Automation of 3d reconstruction of neural tissue from large volume of conventional serial section transmission electron micrographs. J. Neurosci. Methods 176, 276–289 (2009)
6. Reina, A.V., Miller, E., Pfister, H.: Multiphase geometric couplings for the segmentation of neural processes. In: CVPR (June 2009)
7. Kannan, U., Paiva, A., Jurrus, E., Tasdizen, T.: Automatic markup of neural cell membranes using boosted decision stumps. In: ISBI (July 2009)
8. Jurrus, E., Whitaker, R., Jones, B.W., Marc, R., Tasdizen, T.: An optimal-path approach for neural circuit reconstruction. In: ISBI, pp. 1609–1612 (2008)
9. Jurrus, E., Tasdizen, T., Koshevoy, P., Fletcher, P.T., Hardy, M., Chien, C., Denk, W., Whitaker, R.: Axon tracking in serial block-face scanning electron microscopy. Med. Image Anal. 13(1), 180–188 (2009)
10. Breiman, L.: Random forests. Mach. Learn. 45(1), 5–32 (2001)
11. Chui, H., Rangarajan, A.: A new algorithm for non-rigid point matching. In: CVPR, vol. 2, pp. 44–51 (2000)
12. Turaga, S.C., Murray, J.F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W., Seung, H.S.: Convolutional networks can learn to generate affinity graphs for image segmentation. Neural Comput. 22(2), 511–538 (2010)

Analysis of the Striato-Thalamo-Cortical Connectivity on the Cortical Surface to Infer Biomarkers of Huntington's Disease*

Linda Marrakchi-Kacem[1,2], Christine Delmaire[2,4], Alan Tucholka[1,2,3], Pauline Roca[1,2], Pamela Guevara[1,2], Fabrice Poupon[1,2], Jérôme Yelnik[5], Alexandra Durr[5], Jean-François Mangin[1,2], Stéphane Lehericy[2,4], and Cyril Poupon[1,2]

[1] Neurospin, CEA, Gif-sur-Yvette, France
[2] Institut Fédératif de Recherche 49, Gif-sur-Yvette, France
[3] Parietal Team, INRIA Saclay-Ile-de-France, Saclay, France
[4] CENIR, Pitié Salpêtriére Hospital, Paris, France
[5] CRICM, Inserm/UPMC, UMR-S 975, Paris, France

Abstract. The deep brain nuclei play an important role in many brain functions and particularly motor control. Damage to these structures result in movement disorders such as in Parkinson's disease or Huntington's disease, or behavioural disorders such as Tourette syndrome. In this paper, we propose to study the connectivity profile of the deep nuclei to the motor, associative or limbic areas and we introduce a novel tool to build a probabilistic atlas of these connections to the cortex directly on the surface of the cortical mantel, as it corresponds to the space of functional interest. The tool is then applied on two populations of healthy volunteers and patients suffering from severe Huntington's disease to produce two surface atlases of the connectivity of the basal ganglia to the cortical areas. Finally, robust statistics are used to characterize the differences of that connectivity between the two populations, providing new connectivity-based biomarkers of the pathology.

Keywords: deep nuclei, diffusion, tractography, Q-Ball imaging, connectivity, surface atlases.

1 Introduction

The deep brain nuclei include the basal ganglia and the thalamus. The basal ganglia include the caudate nuclei and the putamen, which constitute the striatum, and the globus pallidus. The striatum receives afferents from the cerebral cortex. Most areas of the neocortex except the primary visual and auditive areas have projections on to the striatum. The striatum projects to the external and

* This work was funded by the Association France Parkinson, the Ecole des Neurosciences Paris-Ile-de-France, the ANR MNP, the DHOS AP-HP INSERM and the CHDI High Q Foundation.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 217–224, 2010.

internal segments of the globus pallidus and the internal globus pallidus projects in turn to the thalamus. In Huntington's disease, these pathways are disrupted. Imaging can provide biomarkers that may improve the understanding of the pathophysiology of the disease. Diffusion-weighted (DW) magnetic resonance imaging (dMRI) is now a well established technique to infer the anatomical connectivity *in vivo*. dMRI can probe the anisotropy of the displacement of water molecules at microscopic scales in tissues, revealing their structural organization. In the case of the brain white matter (WM), the more recent local mathematical models of the diffusion process proposed in the literature clearly depict some displacement (PDF) or orientational probability distribution (ODF), the maxima of which corresponds to the direction of the underlying axonal fibers. Tractography algorithms were developed to more or less robustly recover the global connectome using this local information and are now widely used to perform studies of the anatomical connectivity and of its disorders.

Several studies have been conducted about the striato-pallido-thalamo-cortical connectivity [1,2,3]. The first preclinical studies have focused on the functional role played by the thalamus receiving afferents from the cortex and sending projections back to it. Then, surprisingly, most of the clinical studies focused on parcelating the functional territories of the deep structures having the a priori knowledge of their connectivity to the cortical mantel and a labeling of the cortex into functional areas [4,5], whereas none of them has tried to infer the connectivity profile of the deep nuclei (and possibly their sub-territories) onto the cortical mantel. We believe that this approach may be of interest as the deep nuclei are connected to most of the cortical regions. As shown previously, many motor disorders may be linked to putative disruption of the neuronal pathways between specific nuclei and cortical areas. Therefore, building statistical atlases of the connectivity of these nuclei to the cortical mantel for different populations (healthy volunteers versus patients) may help understanding which cortical areas are significantly "disconnected" for each deep structure from both a qualitative and quantitative point of view.

In this paper we propose a novel tool to study the striato-thalamo-cortical connectivity at the surface of the cortical mantel, dedicated to group studies. This tool relies on the construction of surface probabilistic atlas that can be used to quantify the probability of connection of a given nucleus to predefined cortical regions of interest for a given subject or for a given population. We will show that the measure of the probabilities provided by this tool can be used to detect and characterize the axonal disruptions occurring in Huntington's disease. The paper is organized as follows: after introducing the methods involving several image processing steps to project the information onto the cortical surface, the technique is applied to two populations of controls and patients; robust statistics are then performed to detect the functional areas of the cortex that were significantly atrophiated. We show that this tool brings complementary information to the previous approaches that have focused on the study of the Huntington's disease (HD) using dMRI [6,7,8].

2 Methods

The method is based on the construction of probabilistic atlases at the surface of the cortical mantel of the connections between the deep nuclei and the cortical areas. This task relies on the use of both T_1-weighted data to extract the deep structures and the cortical mantel and on high angular resolution diffusion-weighted data (HARDI) to recover the anatomical connectivity. This section is devoted to the description of the different steps of image processing required to build the atlases and to analyze them using adequate statistics.

2.1 Structural Database

A database of 17 healthy volunteers and 17 HD patients was acquired on a Tim Trio 3T MRI system (Siemens, Erlangen) in the frame of a clinical project dedicated to the study of Huntington's disease (HD, HDTrack project). All subjects signed an informed consent and the imaging protocol was approved by the Local Ethical Committee. T_1-weighted and DW data were acquired using the following sequence parameters: *3D MPRAGE* TE/TR=2.98ms/2.3s, FOV=256mm, matrix 256×256, TH=1.1mm, 160 slices per slab, RBW=240Hz/pixel; *Single-shot twice refocused spin-echo DW-EPI* TE/TR=86ms/12s, FOV= 256mm, matrix 128x128, TH=2mm, 80 slices, acceleration factor 2, partial Fourier 6/8, RBW=1630Hz/pixel, b=1000s/mm^2, 50 diffusion directions uniformly distributed. EPI distortions caused by susceptibility effects were corrected with BrainVISA and using a further phase map acquisition to evaluate the distortions along the phase axis. Each DW data was matched to the corresponding T_1-weighted data using a rigid 3D transform.

2.2 Segmentation of the Deep Nuclei and the Cortex

The deep nuclei and cortex were segmented from the T_1-weighted data. For each subject, eight deep structures were automatically delineated using the method described in [9]: left caudate (LCd), left putamen (LPu), left thalamus (LTh), left globus pallidus (LGP), right caudate (RCd), right putamen (RPu), right globus pallidus (RGP), right thalamus (RTh). The automatic segmentations were then checked by a neuro-anatomist and corrected manually if needed.

FreeSurfer was used to extract the interface between the white matter (WM) and the cortex for all the subjects. The vertices of the obtained surface meshes are in direct correspondence [10]. This property is mandatory to conveniently match the different subjects [11]. In order to synthetize the information stemming from different subjects, an average surface was computed for the two populations from the individual interfaces extracted for all the subjects.

2.3 Inference of the Striato-Pallido-Thalamo-Cortical Connectivity

Tractography. In order to infer the connectivity between the deep nuclei and the cortical surface, a streamline probabilistic tractography algorithm was

employed as described in [12]. This choice was motivated by efficacy of such algorithms to deal with complex fiber geometries (crossings, fannings) compared to streamline deterministic algorithms. Bayesian or global algorithms could have been chosen, but at the price of much longer computation times. A robust mask of the brain white matter was built from the T_1-weighted data and 10 probabilistic streamlines were processed for each voxel of the mask. An analytical Q-ball model described in [13] was used to estimate the local underlying orientation distribution function (ODF) using a spherical harmonics order 6 and a regularization factor equal to 0.006. At each step of the streamlining, the most likely direction is determined from the ODF and a random direction is chosen in a cone of aperture 30° around the optimal direction. This whole brain tractography leads to individual tractograms containing 10^6 fibers on average.

The intersection of each fiber with the deep nuclei was computed. Starting from one of the two extremities of a fiber, if n is the first nucleus to be met by the points of the fiber, the portion of the fiber linking the chosen fiber extremity to n is attributed to n. This process provides the fibers crossing each nucleus.

Striato-Pallido-Thalamo-Cortical Connectivity Matrix. A connectivity matrix was computed to evaluate at each vertex of the cortical surface the number of fibers linking it to each nucleus. For each nucleus n, the number of fibers connecting it to each region of the cortex is obtained by computing the intersection between the fibers crossing n and the WM/cortex interface. The values related to each nucleus n are stored in a line of a sparse matrix C_s^{raw} for subject s. This matrix is then smoothed over the surface to account for a reasonable uncertainty on the tracking result giving a matrix C_s for the subject s [14].

2.4 Probabilistic Surface Atlases of the Connectivity

The goal is now to merge all the striato-thalamo-pallido-cortical connectivity information stemming from all the subjects into a single probabilistic atlas. This task must be repeated for the two different populations.

For any subject s, its connectivity matrix contains for each nucleus n and for any vertex position v an approximation $C_s(n, v)$ of the number of fibers connecting n to v. The probability $p_s(n, v)$ of connection between n and v is the ratio between $C_s(n, v)$ and the global number of fibers coming from all the nuclei that project onto the cortical mantel of the subject s. Let $\mathcal{N}$ be the set of nuclei and $\mathcal{V}_s$ be the set of positions of the vertices belonging to the WM/cortex interface for the subject s. For any $n \in \mathcal{N}$ and any $v \in \mathcal{V}_s$,

$$p_s(n, v) = \frac{C_s(n, v)}{\sum_{n' \in \mathcal{N}} \sum_{v' \in \mathcal{V}_s} C_s(n', v')} \tag{1}$$

For a given nucleus n, the set of values $p_s(n, v)$ associated to the vertices belonging to the WM/cortex interface represents the probabilistic atlas of the connectivity of this nucleus n for the subject s. For a given population $\mathcal{P}$, averaging the connectivity matrices C_s provides the wanted probabilistic atlas for all the

nuclei n, since vertices of the individual surfaces provided by FreeSurfer are in direct correspondence. $(\forall s, \forall s', \mathcal{V}_s = \mathcal{V}_{s'} = \mathcal{V})$.

$$C_{\mathcal{P}}(n, v) = \frac{\sum_{s \in \mathcal{P}} C_s(n, v)}{card(\mathcal{P})} \tag{2}$$

The probability $p_{\mathcal{P}}(n, v)$ of connection of $n \in \mathcal{N}$ to $v \in \mathcal{V}$ is the ratio between $C_{\mathcal{P}}(n, v)$ and the mean total number of fibers coming from all the nuclei that project onto the average cortical mantel for the population $\mathcal{P}$. For a given nucleus n, the set of values $p_{\mathcal{P}}(n, v)$ computed over all the vertices belonging to the WM/cortex average interface of $\mathcal{P}$ represents the probabilistic atlas of the connectivity of n, for the population $\mathcal{P}$.

$$p_{\mathcal{P}}(n, v) = \frac{C_{\mathcal{P}}(n, v)}{\sum_{n' \in \mathcal{N}} \sum_{v' \in \mathcal{V}} C_{\mathcal{P}}(n', v')} \tag{3}$$

Probabilistic atlases have been represented for each nucleus using a gradient color ranging from red corresponding to a high probability of connection to gray corresponding to a low probability of connection. Thus, simple qualitative comparisons can already be performed between the two populations of healthy subjects and patients.

2.5 Atlas Differences between Populations

The HDTrack database includes 2 populations: healthy subjects (H) and subjects suffering from severe Huntington disease (HD). Probabilistic connectivity atlases computed, for each nucleus, on each subject of each population can be used to perform some statistical comparisons between the two populations.

A Mann Whitney test was used to compare the probability of connection of each nucleus to the cortex between the two populations. This choice was motivated by the lack of assumption about the normality of the data.

For a given population $\mathcal{P}$, the probability $p_s(n, v)$ was considered as a random variable $X^{\mathcal{P}}_{n,v}$ taking different values for the different subjects s. Using the Mann-Whitney test on the two random variables $X^{\mathcal{P}_{\mathcal{H}}}_{n,v}$ and $X^{\mathcal{P}_{\mathcal{HD}}}_{n,v}$ corresponding to the two populations of controls and patients, we detected trends of significant difference (p-value< 0.05) of the connectivity, for the given nucleus n and the given vertex position v, between the two populations $\mathcal{P}_{\mathcal{H}}$ and $\mathcal{P}_{\mathcal{HD}}$.

3 Results and Discussion

Probabilistic Connectivity Atlases. A probabilistic connectivity atlas was computed for the two populations (H) and (HD), and for all the nuclei, as depicted in figure 1. The left and right columns correspond to the left and right structures respectively. The intensity of the red color is proportional to the level of connectivity. We can observe that most of the connections of a given nucleus are located in the ipsilateral hemisphere. The globus pallidus does not present

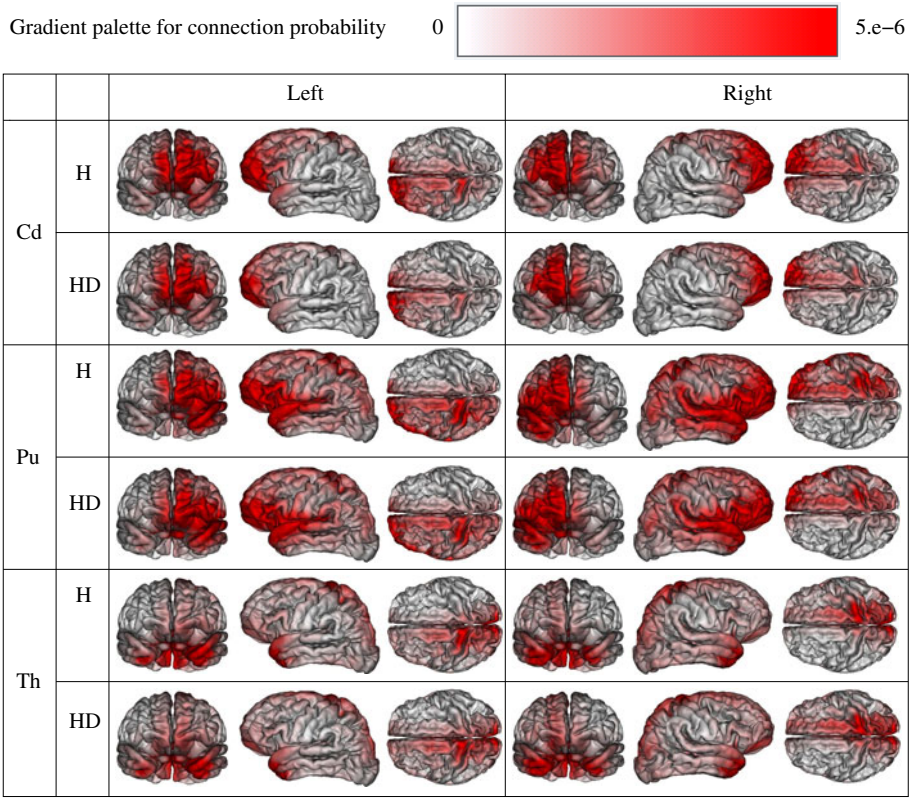


Fig. 1. Surface probabilistic striato-thalamo-cortical connectivity atlases computed for each population: healthy subjects and patients suffering from Huntington's disease

any projection to the cortex and is consequently not shown in the figure. As expected, the caudate nuclei project mostly to the frontal areas and to the premotor and motor areas. The putamen is mainly connected to the motor and premotor areas, to the frontal ventro-lateral areas, and to the temporal superior area. The thalami have projections to the entire cortex.

Statistical Comparison Test. For each nucleus, the Mann Whitney test allowed the detection of the vertices presenting a significant connectivity difference between the two populations. The regions corresponding to these vertices were represented using a color palette as shown in figure 2.

Reduction of the connections of the caudate nucleus was greater than the reduction of the connections of the putamen and thalami, especially in the Broca's area, the motor area, and the temporal superior area. The connectivity differences observed obtained by the comparison of the probabilistic connectivity atlases were in good agreement with the symptoms observed on the HD patients. As an example, the probability of connection of the left caudate, left putamen,

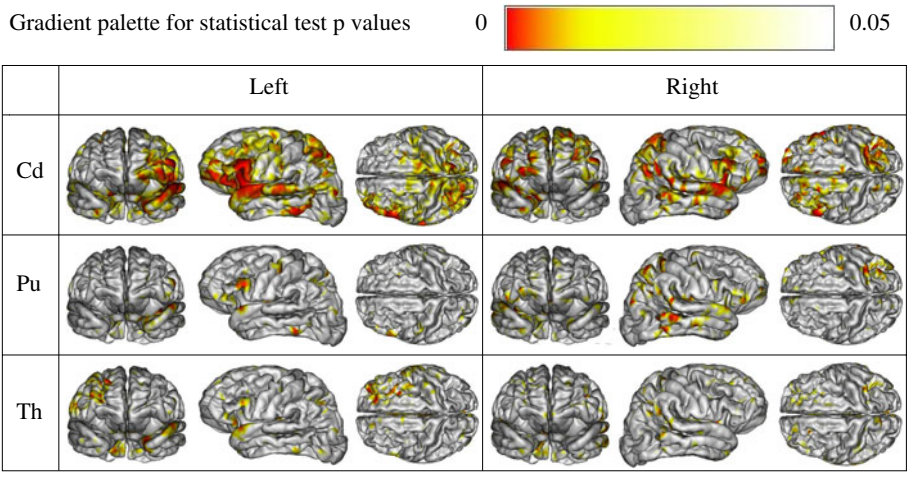


Fig. 2. Cortical regions presenting a significant connectivity difference between healthy subjects and Huntington's patients per nucleus stemming from statistical analysis of their striato-thalamo-cortical surface atlases

left thalamus and right caudate to the Broca's area was significantly different between the two populations. The obtained statistics clearly indicate that the caudate nucleus has the most atophiated connectivity and this atrophy is more predominant in the left caudate nucleus.

3.1 Conclusion

In this paper, we introduced a novel tool for the study of the striato-thalamo-cortical connectivity relying on surface probabilistic connectivity atlases. Projecting the connectivity profile onto the cortical mantel is relevant as it enables to detect disruption of these connections directly in the frame of functional areas. We used this novel tool on a population of healthy subject and on a population of patients suffering of the Huntington's disease and we proved that it was adequate to detect differences of the striato-thalamo-cortical connectivity between the two populations, using robust statistics. The structural lesions that were detected were in good agreement with the known physiopathology of Huntington's disease. In the future, this tool will be used to investigate several neurodegenerative pathologies involving the basal ganglia, and improvements will be done to provide accurate information about areas with atrophy, such as performing longitudinal studies of the pathology.

References

1. Alexander, G.E., DeLong, M.R., Strick, P.L.: Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annual Review of Neuroscience 9(1), 357–381 (1986)

2. Barbas, H., Pandya, D.: Architecture and frontal cortical connections of the premotor cortex (area 6) in the rhesus monkey. The Journal of Comparative Neurology 256, 211–228 (1987)
3. Alexander, G.E., Crutcher, M.D.: Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends in Neurosciences 13(7), 266–271 (1990)
4. Behrens, T., Johansen-Berg, H., Woolrich, M.W., Smith, S.M., Wheeler-Kingshott, C.A.M., Boulby, P.A., Barker, G.J., Sillery, E.L., Sheehan, K., Ciccarelli, O., Thompson, A.J., Brady, J.M., Matthews, P.M.: Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Neuroscience 6, 750–757 (2003)
5. Draganski, B., Kherif, G., Kloppel, S., Cook, P.A., Alexander, D.C., Parker, G.J., Deichmann, R., Ashburner, J., Frackowiak, R.S.: Evidence for segregated and integrative connectivity patterns in the human basal ganglia. The Journal of Neuroscience 28, 7143–7152 (2008)
6. Bohanna, I., Georgiou-Karistianis, N., Hannan, A.J., Egan, G.F.: Magnetic resonance imaging as an approach towards identifying neuropathological biomarkers for huntington's disease. Brain Res. Rev. 58(1), 209–225 (2008)
7. Douaud, G., Behrens, T.E., Poupon, C., Cointepas, Y., Jbabdi, S., Gaura, V., Golestani, N., Krystkowiak, P., Verny, C., Damier, P., Bachoud-Levi, A.C., Hantraye, P., Remy, P.: In vivo evidence for the selective subcortical degeneration in huntington's disease. Neuroimage 46, 958–966 (2009)
8. Rosas, H.D., Lee, S.Y., Bender, A.C., Zaleta, A.K., Vangel, M., Yu, P., Fischl, B., Pappu, V., Onorato, C., Cha, J.H., Salat, D.H., Hersch, S.M.: Altered white matter microstructure in the corpus callosum in huntington's disease: Implications for cortical disconnection. NeuroImage 49(4), 2995–3004 (2010)
9. Marrakchi-Kacem, L., Poupon, C., Mangin, J.F., Poupon, F.: Multi-contrast deep nuclei segmentation using a probabilistic atlas. In: ISBI (2010)
10. Argall, B.D., Saad, Z.S., Beauchamp, M.S.: Simplified intersubject averaging on the cortical surface using suma. Human Brain Mapping 27, 14–27 (2006)
11. Fischl, B., Sereno, M.I., Tootell, R.B., Dale, A.M.: High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping 8, 272–284 (1999)
12. Perrin, M., Poupon, C., Cointepas, Y., Rieul, B., Golestani, N., Pallier, C., Riviere, D., Constantinesco, A., LeBihan, D., Mangin, J.F.: Fiber tracking in q-ball fields using regularized particle trajectories. In: Christensen, G.E., Sonka, M. (eds.) IPMI 2005. LNCS, vol. 3565, pp. 52–63. Springer, Heidelberg (2005)
13. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast, and robust analytical q-ball imaging. Magnetic Resonance in Medicine 58(3), 497–510 (2007)
14. Roca, P., Riviere, D., Guevara, P., Poupon, C., Mangin, J.F.: Tractography-based parcellation of the cortex using a spatially-informed dimension reduction of the connectivity matrix. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 935–942. Springer, Heidelberg (2009)

The Fiber Laterality Histogram: A New Way to Measure White Matter Asymmetry

Lauren J. O'Donnell[1], Carl-Fredrik Westin[1], Isaiah Norton[1], Stephen Whalen[1], Laura Rigolo[1], Ruth Propper[2], and Alexandra J. Golby[1]

[1] Brigham and Women's Hospital, Harvard Medical School, Boston MA, USA
odonnell@bwh.harvard.edu
[2] Merrimack College, Psychology Department, North Andover, MA, USA*

Abstract. The quantification of brain asymmetries may provide biomarkers for presurgical localization of language function and can improve our understanding of neural structure-function relationships in health and disease. We propose a new method for studying the asymmetry of the white matter tracts in the entire brain, and we apply it to a preliminary study of normal subjects across the handedness spectrum. Methods for quantifying white matter asymmetry using diffusion MRI tractography have thus far been based on comparing numbers of fibers or volumes of a single fiber tract across hemispheres. We propose a generalization of such methods, where the "number of fibers" laterality measurement is extended to the entire brain using a soft fiber comparison metric. We summarize the distribution of fiber laterality indices over the whole brain in a histogram, and we measure properties of the distribution such as its skewness, median, and inter-quartile range. The whole-brain fiber laterality histogram can be measured in an exploratory fashion without hypothesizing asymmetries only in particular structures. We demonstrate an overall difference in white matter asymmetry in consistent- and inconsistent-handers: the skewness of the fiber laterality histogram is significantly different across handedness groups.

1 Introduction

Brain asymmetries provide clues about the brain's functional organization. For example, known left-greater-than-right perisylvian *asymmetries* [1] relate to the localization of language function to the left hemisphere in most right-handed subjects. But interestingly, increased *symmetry* of the direct segment of the arcuate fasciculus fiber tract has been shown to relate to improved verbal recall performance [2]. These seemingly contradictory results indicate that much remains to be learned about how morphological asymmetries may underlie major functional differences across the hemispheres. Because the study of fiber tract asymmetry using imaging is a relatively recent field, the development of new methods to

* We acknowledge the following grant support: NIH U41RR019703, R01MH074794, R25CA089017, P01CA067165, P41RR013218, Brain Science Foundation, and Klarmin Family Foundation.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 225–232, 2010.

measure white matter (WM) asymmetry using diffusion tensor MRI (DTI) may provide useful tools for studies of structure and function in healthy and diseased subjects. Current approaches for assessing WM asymmetries include voxel-based methods with normalization to a symmetric template and tractography-based methods where fiber counts, volumes, or FA values are measured.

Brain asymmetries have been extensively studied in healthy *right-handed subjects*, where the most prominent asymmetries are related to language areas such as the planum temporale, and to the finding that the frontal lobe is larger on the right and the occipital lobe is larger on the left (termed petalia and brain torque) [1]. These asymmetries have been measured (for example) with voxel-based approaches in structural MRI [3]. Thus far DTI-based methods have mainly demonstrated white matter asymmetries in the arcuate fasciculus (AF). A voxel-based statistical analysis using a symmetric template found strong asymmetries of FA in the arcuate fasciculus (higher FA in the left hemisphere) in consistent right-handers [4] (though this was not found in another similar study [5]). Despite different anatomical subdivisions of the AF (superior temporal and middle temporal connections, vs. direct and indirect segments) studies of DTI tractography in right handers have robustly found greater left-vs-right hemisphere AF volume [6], and fiber trajectory counts [2]. For the direct segment of AF, 62.5% of right handers have complete left lateralization and 17.5% have a symmetric AF [2]. Motor-related asymmetries were also shown with more extensive tractography connectivity to motor cortex in the left hemisphere [7], and higher FA in left hemisphere tractography [8]. But a different study did not find corticospinal tract asymmetry in terms of number of fibers [9].

The *relationship of brain asymmetry to handedness* is more subtle. Voxel-based methods using structural MRI have found no effect of handedness and no handedness interaction with asymmetry [3]. However, cortical and DTI methods have found asymmetries related to motor regions. Central sulcus depth was deeper in the left hemisphere in male consistent right handers [10], and this pattern may be reversed in left handers but it did not reach significance in the group (of 465 subjects). With DTI a reversed pattern in left- and right-handers was found in a much smaller cohort of 28 subjects: higher FA was detected in the precentral gyrus in the hemisphere contralateral to the dominant hand [4]. Language asymmetries have also been studied in relation to handedness using DTI tractography, with mixed results so far. One study found no effect of handedness, with leftward asymmetry of AF tractography regardless of handedness or functional language lateralization [11], whereas another study found AF asymmetry related to handedness in men but not women [12].

On the basis of these results from the literature it appears that DTI, in comparison with structural MRI, may be more sensitive to subtle brain differences related to handedness. However, results have been inconsistent across studies, and much of the brain's white matter has not been systematically explored. These facts inspire our investigation of a hypothesis-free fiber asymmetry measurement method.

2 Methods

We propose a generalization of DTI tractography fiber counting or fiber volume measurement methods, where the "number of fibers" laterality measurement is extended to the entire brain by using a soft fiber comparison. We apply this method to a study of 26 left-, right- and inconsistent-handers.

2.1 Algorithm

The algorithm proposed in this paper can be roughly thought of as follows. Each fiber "looks" in its immediate neighborhood and "counts" the number of similar fibers near it. It also "looks" in the corresponding region in the other hemisphere, and "counts" the number of similar fibers there. The relationship between these two "counts" is expressed as a laterality index (LI) that reflects any asymmetries in the anatomical extent of the traced structure of which the fiber is a part. This process is repeated for each fiber in the brain, so each fiber "votes" for the laterality of the structure it belongs to.

Counting fibers directly is not ideal because it involves hard cutoffs of how similar or nearby a fiber must be to be counted as part of the structure of interest. This is not robust due to variable structure sizes within and across subjects. However, it is possible to use a "soft" count of the number of similar fibers. We propose to determine fiber correspondence using the soft fiber similarity metric between each fiber and all other fibers in its hemisphere, and between the mirror image of that fiber and all the fibers in the other hemisphere. We propose the following *fiber similarity*:

$$\prod_i e^{\frac{-(f_i - n_i)^2}{\sigma^2}} \tag{1}$$

where the ith point on fiber f is f_i and the ith point on another, perhaps neighboring, fiber is n_i, and their distance apart is mapped through a Gaussian kernel to give a number that can be thought of as a local probability of those points being in the same structure. These probabilities are multiplied along both fibers to get an overall probability that the fibers are in the same structure. Sigma (σ) controls the scale of searching for similar fibers: at an extreme σ of 0mm, all fibers will be completely right- or left-lateralized with a LI of +/-1, while at a large σ of say 1000mm all fibers will have a symmetric LI near 0. (We chose σ of 50mm to avoid having large numbers of fibers at +/-1 that gave the histogram a "truncated" appearance, but our results were significant with sigmas from 30-50mm.) The formula relates to similarity metrics used in the fiber clustering field, both within [13] and across hemispheres [14] and was inspired by the interhemispheric comparison of fiber tracts in a beautifully descriptive study of temporal connections [15]. We efficiently implemented the similarity metric using 5 points on each fiber (endpoints, midpoints, and points between them), accounting for point ordering by computing similarities with points in forward and reverse orders then taking the maximum.

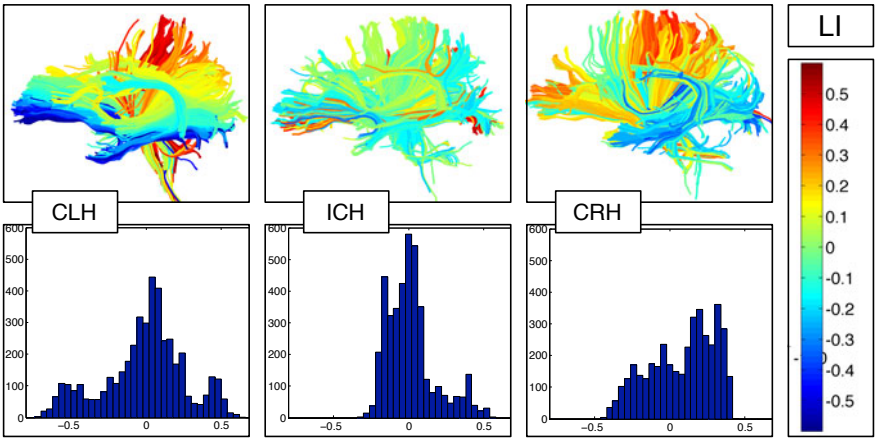


Fig. 1. Fiber laterality indices and fiber laterality histograms in example subjects from each handedness group. Top row: individual subject fibers painted by their LIs, where blue and cyan are left-lateralized, red and yellow are right-lateralized, and green is not lateralized, i.e. it is symmetric. The inconsistent-handed subject, center, has relatively few asymmetric fibers as shown by the prevalence of green color. Bottom row: Fiber laterality histograms showing the distribution of LIs over all the fibers in the brain of each subject from the top row.

To convert the similarities to a *fiber laterality index* for each fiber, we employ a standard laterality index formula that ranges from -1 (left lateralized) to +1 (right lateralized)

$$LI = \frac{R - L}{R + L} \tag{2}$$

where R is the total sum of similarity to all fibers in the right hemisphere and L is the total sum of similarity to all fibers in the left hemisphere. To summarize the distribution of fiber LIs over the whole brain of each subject we construct a *fiber laterality histogram* and we measure the skewness, kurtosis, median, and interquartile range (a measure of the statistical dispersion of the LI data).

2.2 Data and Subjects Studied

Twenty-six individuals participated: 9 men and 17 women (age M=28.54 years, SD=9.19). Participants had no history of neurological problems, psychiatric illness, or head trauma. Handedness was determined via score on the Edinburgh Handedness Inventory [16], where scores can range from -100, indicating perfect consistent left-hand preference, to +100, indicating perfect consistent right-hand preference. Subjects were divided into three handedness groups: consistent left handers (CLH n=5), inconsistent handers (ICH n=16), and consistent right handers (CRH n=5), using a cutoff handedness score of 75 (the median of the absolute value of the handedness scores). There were 5 borderline cases with absolute handedness of ±75 that were placed in the inconsistent group. Our dataset was

specifically gathered to over-represent LH and CLH relative to the population in order to be able to study these groups.

DTI EPI images were acquired at 3T using an 8-channel head coil and ASSET matrix=128x128; FOV=25.6cm; Phase FOV=1.0; B value=1000s/mm2; 55 DWI gradients and 5 baseline T2 images; voxel size=2x2x2.6mm. Whole brain tractography was generated by seeding trajectories (fibers) on a 2mm grid throughout the entire white matter of each subject using Runge-Kutta order two integration in 3D Slicer (www.slicer.org). DTI tractography was normalized to a common coordinate system created by congealing (an entropy-based unbiased group registration method [17]) of all subjects' fractional anisotropy images. Fibers shorter than 75mm, fibers crossing the midsagittal plane, and fibers restricted to the brainstem were discarded before laterality processing.

3 Results

Despite our small sample size, our new method was able to detect significant group differences based on handedness direction (left/right) and handedness degree (consistency/inconsistency). Fiber laterality indices (Figure 1) were successfully measured for all subjects. We tested for differences across groups in the skewness, kurtosis, median, and interquartile range (IQR) of the fiber LIs (using one-way ANOVA). No significant differences were found in median or kurtosis. The IQR was was not significantly different across the three groups ($p = 0.098$) but the measurements (Figure 2, bottom right) indicated a difference between consistent right handers and other groups (other groups mean 0.26, CRH mean 0.36, $p = 0.03$ via t-test) that we believe is related to torque. The skewness was significantly different across the three groups (CLH mean -0.41, ICH mean 0.23, CRH mean -0.35, $p=0.006$) because the ICH group was different from both others (we hypothesize the ICH brains are more symmetric so their LI histograms are less skewed). To ensure our results weren't extremely dependent on the sigma parameter from eq. 1, we tested sigma values of 30, 40, 50, and 60mm. The significant results were reproduced at all scales except 60mm, when significance decreased slightly past the 0.05 threshold. The histograms were truncated at +/-1 for lower sigmas so we chose sigma of 50mm for the results reported here.

4 Discussion

To our knowledge this is the first time the asymmetry of white matter tractography has been quantified in the entire brain. Our results (highly asymmetric tracts as well as brain torque) correspond well with known asymmetries in the highly-studied CRH population. To evaluate the success of our method at detecting asymmetries and to to examine which fiber tracts may be driving the group differences, we selected fiber tracts with very high LIs for viewing (Figure 3), using as input all CRH subjects. The results appear to correspond very well with CRH data from the literature: the left arcuate fasciculus was highly leftward asymmetric (eg. [6,2]). The left cingulum bundle (left > right FA [18]), the

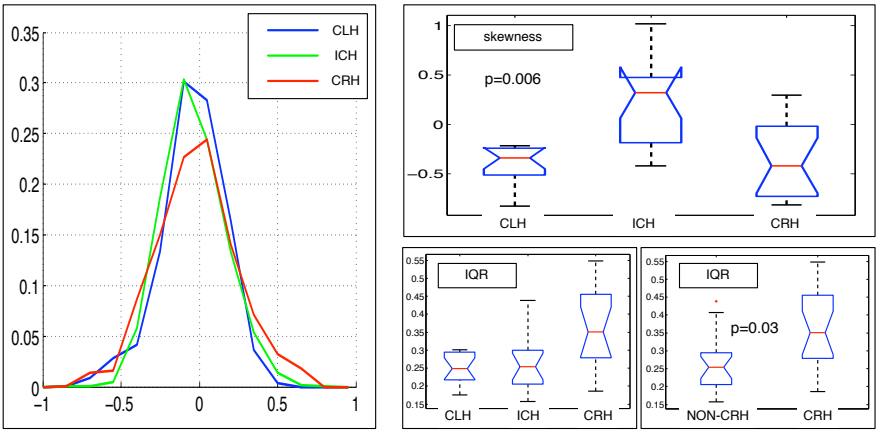


Fig. 2. Differences in the fiber laterality index distribution across groups. Left: average fiber laterality histogram in the three groups (bin counts were normalized within subject by number of fibers, then averaged across subjects in the group at each bin location). Note the tails of the green inconsistent-hander distribution are most symmetric about 0, and the red consistent-right-hander curve appears wider. Right (boxplots): A significant skewness difference was measured between inconsistent handers and the other two groups (top). The interquartile range, bottom, in the three handedness groups was not significantly different ($p = 0.098$), but it did differ if CLH and ICH were combined and compared with consistent right handers.

right anterior indirect segment of the arcuate (right > left FA [2]), and the right uncinate (known to be larger than the left [1]) were highly asymmetric. Additional asymmetric structures were detected whose asymmetry may not have yet been reported. Our method also shows evidence of the large-scale asymmetry of brain torque (Figure 3), and we hypothesize that the the greater IQR we found in right handers corresponds to their known higher brain torque [1]. In our analysis strategy, curve "widening" as seen in the red curve in Figure 2 is indicative of low-level asymmetry as might be seen with brain torque.

Extensions of this study will include quantifying the patterns of asymmetric tracts in the handedness groups and testing the method with the effects of torque removed via registration. One point to mention is that we can't disambiguate the effects of gender and handedness in this small study because 13 of 16 inconsistent handers were female. However, the skewness and IQR were not significantly different across genders.

In summary, we have presented a new way to measure asymmetry in the white matter of the human brain, the *fiber laterality index*, and a new way to compare asymmetry across subjects in the form of a *fiber laterality histogram* and its summary statistics. We applied the method to a small study of brain asymmetry and handedness, detecting significant differences in skewness between consistent- and inconsistent-handed subjects. Our results regarding skewness support the idea that inconsistent-handed people have more symmetric white matter. The sign of skewness is not necessarily expected to reverse for right and left handers (which

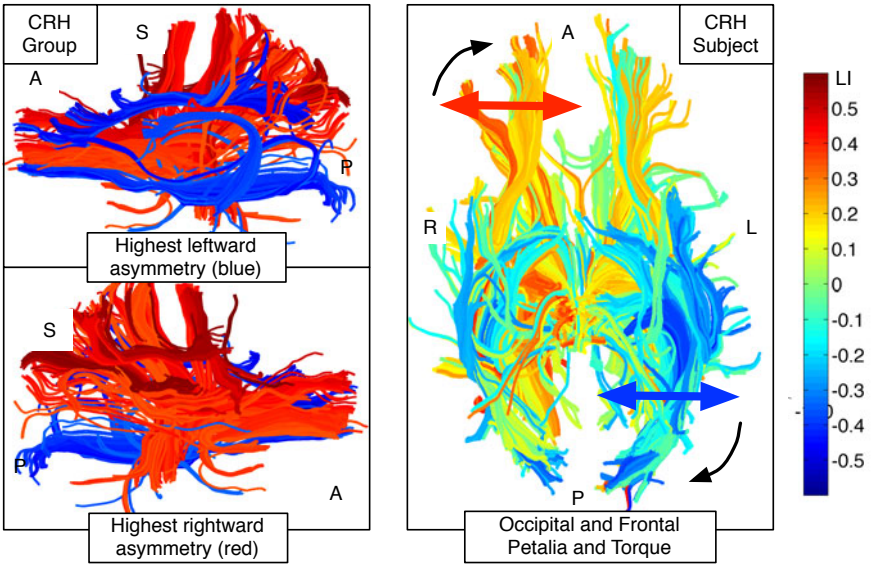


Fig. 3. Left column: Fibers with extreme laterality indices, a composite visualization including fibers from all consistent right handers. Red-blue images show fibers with absolute value of LI greater than 0.45. Notable blue regions (highly left-lateralized in one or more CRH subjects) include the arcuate fasciculus and left cingulum, as well as occipitofrontal connections. Red regions (highly right-lateralized in one or more CRH subjects) include part of anterior thalamic radiation, uncinate, and SLF/anterior indirect segment of the arcuate. Right column: example CRH subject showing left-lateralization of occipital fibers (blue) and right-lateralization of frontal fibers (orange and yellow), potentially corresponding to the known size differences [1] of occipital and frontal lobes in right handers and showing torque of the occipital fibers towards the midline.

would indicate a mirror-image brain in the two groups) because known asymmetries do not frequently reverse, as demonstrated by language function where only some left handers are right-lateralized. So the phenotype of left-handedness does not fully predict the underlying brain organization. Our additional result of greater histogram IQR in right-handers may correspond to their known greater brain torque compared to non-right-handers. Our approach has potential advantages for future investigations of white matter symmetry: (1) whole-brain measurement can be performed without limiting tracts of interest based on a priori hypotheses, and (2) the use of summary statistics from the LI distribution increases statistical power.

References

1. Toga, A., Thompson, P.: Mapping brain asymmetry. Nature Reviews Neuroscience 4(1), 37–48 (2003)
2. Catani, M., Allin, M., Husain, M., Pugliese, L., Mesulam, M., Murray, R., Jones, D.: Symmetries in human brain language pathways correlate with verbal recall. Proceedings of the National Academy of Sciences 104(43), 17163 (2007)

3. Good, C., Johnsrude, I., Ashburner, J., Henson, R., Friston, K., Frackowiak, R.: Cerebral asymmetry and the effects of sex and handedness on brain structure: a voxel-based morphometric analysis of 465 normal adult human brains. Neuroimage 14(3), 685–700 (2001)
4. Buchel, C., Raedler, T., Sommer, M., Sach, M., Weiller, C., Koch, M.: White matter asymmetry in the human brain: a diffusion tensor MRI study. Cerebral Cortex 14(9), 945 (2004)
5. Park, H., Westin, C., Kubicki, M., Maier, S., Niznikiewicz, M., Baer, A., Frumin, M., Kikinis, R., Jolesz, F., McCarley, R., et al.: White matter hemisphere asymmetries in healthy subjects and in schizophrenia: a diffusion tensor MRI study. Neuroimage 23(1), 213–223 (2004)
6. Glasser, M., Rilling, J.: DTI tractography of the human brain's language pathways. Cerebral Cortex 18(11), 2471 (2008)
7. Guye, M., Parker, G., Symms, M., Boulby, P., Wheeler-Kingshott, C., Salek-Haddadi, A., Barker, G., Duncan, J.: Combined functional MRI and tractography to demonstrate the connectivity of the human primary motor cortex in vivo. Neuroimage 19(4), 1349–1360 (2003)
8. Eluvathingal, T., Hasan, K., Kramer, L., Fletcher, J., Ewing-Cobbs, L.: Quantitative diffusion tensor tractography of association and projection fibers in normally developing children and adolescents. Cerebral Cortex 17(12), 2760 (2007)
9. Nucifora, P., Verma, R., Melhem, E., Gur, R., Gur, R.: Leftward asymmetry in relative fiber density of the arcuate fasciculus. Neuroreport 16(8), 791 (2005)
10. Amunts, K., Jancke, L., Mohlberg, H., Steinmetz, H., Zilles, K.: Interhemispheric asymmetry of the human motor cortex related to handedness and gender. Neuropsychologia 38(3), 304–312 (2000)
11. Vernooij, M., Smits, M., Wielopolski, P., Houston, G., Krestin, G., Van der Lugt, A.: Fiber density asymmetry of the arcuate fasciculus in relation to functional hemispheric language lateralization in both right-and left-handed healthy subjects: a combined fMRI and DTI study. Neuroimage 35(3), 1064–1076 (2007)
12. Hagmann, P., Cammoun, L., Martuzzi, R., Maeder, P., Clarke, S., Thiran, J., Meuli, R.: Hand preference and sex shape the architecture of language networks. Human Brain Mapping 27(10), 828 (2006)
13. Wassermann, D., Bloy, L., Kanterakis, E., Verma, R., Deriche, R.: Unsupervised white matter fiber clustering and tract probability map generation: Applications of a Gaussian process framework for white matter fibers. Neuroimage (2010)
14. O'Donnell, L., Westin, C.: Automatic tractography segmentation using a high-dimensional white matter atlas. IEEE Transactions on Medical Imaging 26(11), 1562 (2007)
15. Barrick, T., Lawes, I., Mackay, C., Clark, C.: White matter pathway asymmetry underlies functional lateralization. Cerebral Cortex 17(3), 591 (2007)
16. Oldfield, R.: The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9(1), 97–113 (1971)
17. Zollei, 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)
18. 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(2), 92–98 (2005)

A Geometry-Based Particle Filtering Approach to White Matter Tractography

Peter Savadjiev[1,2], Yogesh Rathi[1], James G. Malcolm[1], Martha E. Shenton[1,2], and Carl-Fredrik Westin[2]

[1] Department of Psychiatry
[2] Department of Radiology
Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

Abstract. We introduce a fibre tractography framework based on a particle filter which estimates a local geometrical model of the underlying white matter tract, formulated as a 'streamline flow' using generalized helicoids. The method is not dependent on the diffusion model, and is applicable to diffusion tensor (DT) data as well as to high angular resolution reconstructions. The geometrical model allows for a robust inference of local tract geometry, which, in the context of the causal filter estimation, guides tractography through regions with partial volume effects. We validate the method on synthetic data and present results on two types *in vivo* data: diffusion tensors and a spherical harmonic reconstruction of the fibre orientation distribution function (fODF).

1 Introduction

Tractography is the process of reconstructing possible white matter fibre pathways from diffusion MRI data. An increasing variety of algorithms dedicated to this problem are available, as reviewed recently in [1,2]. Many such methods are developed with a focus on the specific model used to represent the diffusion process, e.g. [3,4] for diffusion tensors, or e.g. [5,6] for more complex parametric and non-parametric models of diffusion. Additionally, directional information from the diffusion MRI signal is often integrated without a geometrical model of the pathways to be reconstructed. This is the case for existing methods based on the unscented Kalman filter, e.g. [2], or particle filters, e.g. [1,7].

In this paper, we propose a novel tractography approach which models the trajectories of white matter pathways as 3D curves, and infers these curves independently of the underlying model of the diffusion process. The method incorporates a geometrical model for co-varying 3D curve sets, called 'streamline flows' (SF) [8,9], which is infered at each position along the fibre with a particle filter, so that the estimation at each step builds upon previous estimates. With a small parameter set, the SF model captures within a 3D neighborhood $\mathcal{N}$ the full 3D geometry of a collection of curves that may vary within $\mathcal{N}$ in the tangential direction, as well as in the normal and bi-normal directions. Modeling fibre tracts as a collection of curves that 'flow' together in 3D space, and infering a local geometrical model for such structures allows for increased robustness in

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 233–240, 2010.

tractography to partial volume effects in areas where different fibre populations with distinct orientations co-exist. This is particularly useful when working with DT data, as tractography may then proceed to areas not reachable with standard methods. We demonstrate our method with synthetic and *in vivo* data, with diffusion tensors as well as a spherical harmonic reconstruction of the fODF.

2 3D Streamline Flows and the Generalized Helicoid

The work in [8,9] argues for the representation of white matter fibres as sets of dense, locally parallel 3D curves called 'streamline flows', and derives their differential geometry, which is characterized by three curvature functions: the *tangential*, *normal* and *bi-normal* curvatures. A local model for such flows is then proposed in [8,9] which consists of two orientation functions $\theta(x, y, z)$ and $\phi(x, y, z)$, which define the local orientation of the flow (its tangent vector) at every point (x, y, z) in E^3 (three-dimensional Euclidean space):

$$\begin{aligned} \theta(x, y, z) &= \tan^{-1}\left(\frac{K_T x + K_N y}{1 + K_N x - K_T y}\right) + K_B z \ , \\ \phi(x, y, z) &= \alpha\theta(x, y, z) \ . \end{aligned} \tag{1}$$

Here α is a constant, and K_T, K_N and K_B are scalar parameters that specify the values of the tangential, normal and bi-normal curvatures of the flow. This formulation is justified in [8,9] via minimal surface theory as a smooth local model for 3D streamline flows with a small parameter set that describes the flow geometry. In fact, the formulation for θ (and ϕ) given in (1) is that of a *generalized helicoid*. Generalized helicoids are minimal hypersurfaces in n-dimensional Euclidean space E^n, whose geometry has been studied in e.g. [10].

3 Particle Filter Based Tractography

3.1 Theory

This section reviews the particle filter method, described in full elsewhere [11,12]. Details specific to our implementation are given in Section 3.2.

Let $S_t \in \mathbf{R}^n$ be a state vector at time t evolving at discrete time steps according to $S_{t+1} = f_t(S_t) + n_t$, where n_t is i.i.d. random noise with known probability distribution function (pdf). At time t, observations $Y_t \in \mathbf{R}^p$ become available. These measurements relate to the state vector via the observation equation $Y_t = h_t(S_t) + \nu_t$, where ν_t is measurement noise with known pdf. It is assumed that the initial state distribution $p(S_0)$, the state transition function denoted by f_t and the observation likelihood given the state, denoted by $p(Y_t|S_t)$, are known. The particle filter is a sequential Monte Carlo method which produces at each time t a cloud of K particles, $\{S_t^{(i)}\}_{i=1}^K$, whose empirical measure follows closely $p(S_t|Y_{1:t})$, the posterior distribution of the state given past observations.

The algorithm starts by drawing K samples (particles) from the initial state distribution $p(S_0)$ in order to approximate it by $p_0^K(S) = \frac{1}{K}\sum_{i=1}^K \delta(S_0 - S_0^{(i)})$,

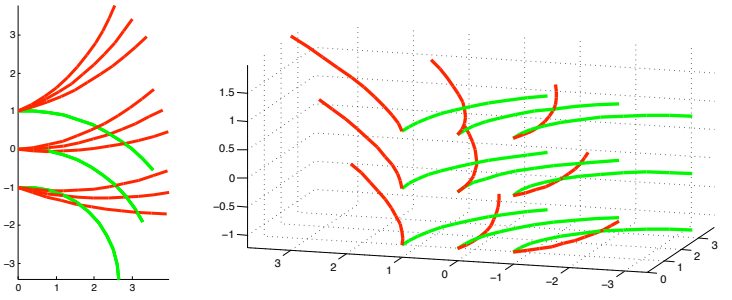


Fig. 1. Two views of a set of two streamline flow examples (one in green and one in red) generated using (1), each sampled with nine streamlines. Left: An orthographic projection from above. Right: A 3D view. The flow parameters are as follows: Green: $K_T = -0.3$, $K_N = 0$, $K_B = 0$, $\alpha = 0$. Red: $K_T = 0.2$, $K_N = 0.3$, $K_B = 0.1$, $\alpha = 0.3$.

and then implements Bayes' recursion at each time step. Assuming that one can sample from the posterior distribution $p(S_t|Y_{1:t})$, an empirical estimate of this distribution is given by $p_t^K(S_t|Y_{1:t}) = \sum_{i=1}^{K} \omega_t^{(i)} \delta(S_t - S_t^{(i)})$, where $\omega_t^{(i)}$ is the weight associated with the i-th particle. In this paper, the state process is assumed to be Markovian, i.e., the likelihood can be expressed as $p(Y_t|S_{0:t}) = p(Y_t|S_t)$. This leads to the following recursion relation for the weights [12]:

$$\omega_t^{(i)} \propto \omega_{t-1}^{(i)} \frac{p(Y_t|S_t^{(i)}) p(S_t^{(i)}|S_{t-1}^{(i)})}{q(S_t^{(i)}|S_{t-1}^{(i)}, Y_t)}, \tag{2}$$

where $q(\cdot)$ is the importance sampling density, and $p(S_t^{(i)}|S_{t-1}^{(i)})$ is the prior distribution that ensures smoothness. The empirical distribution for the posterior is then given by $p_t^K(S_t|Y_t) = \sum_{i=1}^{K} \tilde{\omega}_t^{(i)} \delta(S_t - S_t^{(i)})$, where $\tilde{\omega}_t^{(i)}$ are normalized weights. Finally, resampling is performed so that particles with low weights are eliminated [12]. At each time step, the maximum *a posteriori* (MAP) estimate S_t^{MAP} given by the particle with the largest normalized weight $\tilde{\omega}_t^{MAP}$ is stored and used for tractography as described in Section 3.3.

3.2 Implementation

Our implementation uses a state vector $S = \{K_T, K_N, K_B, \alpha, m_x, m_y, m_z, \beta\}$, where the first four components parametrize a local SF model for the underlying tracts, as defined by (1). The last four determine the orientation of the SF model in E^3, with vector $\mathbf{m} = [m_x, m_y, m_z]$ and a rotation angle β around $\mathbf{m}$. Two SF examples generated using (1) are visualized in Fig. 1.

The initial state distribution $p(S_0)$ is obtained by setting the m_x, m_y and m_z components of each particle to the principal eigenvector of a DT reconstruction in each seed voxel. The other five state components are set to 0. The state transition function f_t is then used to propagate the particles with normally distributed noise $n_t \sim N(0, \Sigma)$, where Σ is user-defined.

Algorithm 1. The particle filter based tractography algorithm.

foreach *Seed Point* **do**
  Compute $p(S_0)$ as described in §3.2;
  t $\leftarrow 1$;
  while *(G)FA > threshold* **and** *current tract position is inside brain mask* **do**
    for $i = 1$ **to** K **do**
      $S_t^{(i)} \leftarrow S_{t-1}^{(i)} + n_t$;
      Compute $\omega_t^{(i)}$ using (2), as described in §3.2;
    $\tilde{\omega}_t^{(i)} \leftarrow \omega_t^{(i)} / \sum_i^K \omega_t^{(i)}$;
    $\tilde{\omega}_t^{MAP} \leftarrow \max_i(\tilde{\omega}_t^{(i)})$;
    Perform local tractography using S_t^{MAP}, as described in §3.3;
    Perform particle resampling [12];
    $t \leftarrow t + 1$;

The observation likelihood $p_t(Y_t|S_t^{(i)})$ is computed by instantiating the streamline flow defined by $S_t^{(i)}$ in a voxel neighborhood $\mathcal{N}$ of size $n \times n \times n$. At each voxel with position vector $\mathbf{x} = (x, y, z)$ in $\mathcal{N}$, the ODF $\Psi_{\mathbf{x}}$ is then evaluated in the direction of the streamline flow at $\mathbf{x}$, as given by (1). The observation likelihood is then defined as the average of these values:

$$p_t(Y_t|S_t^{(i)}) \equiv \frac{1}{|\mathcal{N}|} \sum_{\mathbf{x} \in \mathcal{N}} \Psi_{\mathbf{x}}(\theta(\mathbf{x}), \phi(\mathbf{x})) \quad . \tag{3}$$

Note that no restrictions are placed on Ψ, as it may be a diffusion tensor, or any other diffusion or fibre ODF.

To ensure smoothness of the tractography, we define a prior based on S_{t-1}^{MAP} rather than on $S_{t-1}^{(i)}$, i.e., $p(S_t^{(i)}|S_{t-1}^{(i)}) \propto \exp(-\|S_t^{(i)} - S_{t-1}^{MAP}\|/\sigma)$. Finally, we set $q(\cdot)$ to $q(\cdot) \sim N(0, \Sigma)$. Both σ and Σ are user-defined.

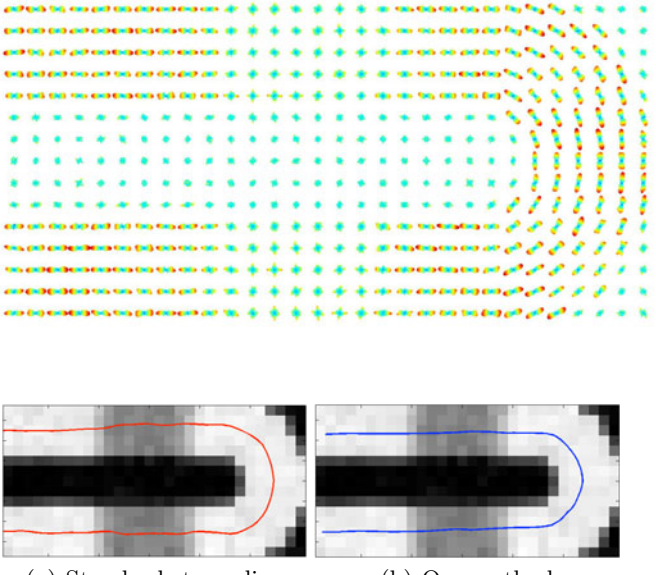
3.3 Tractography Algorithm

Given the MAP estimate S_t^{MAP} obtained at each time step, our method consists in performing streamline tractography over a short distance following the SF field defined by S_t^{MAP}. Given a neighborhood $\mathcal{N}$ of size $n \times n \times n$ voxels, we follow the streamline flow over a distance of $n/3$ voxels starting from the center of $\mathcal{N}$. Performing tractography by following the infered streamline flow model allows for a regularization of the path and for robustness to partial volume effects.

4 Experiments and Results

4.1 Synthetic Data Validation

A synthetic fODF volume that simulates a U-fibre with two 90° crossing regions was generated using the methodology described in [13]. First, a diffusion-weighted



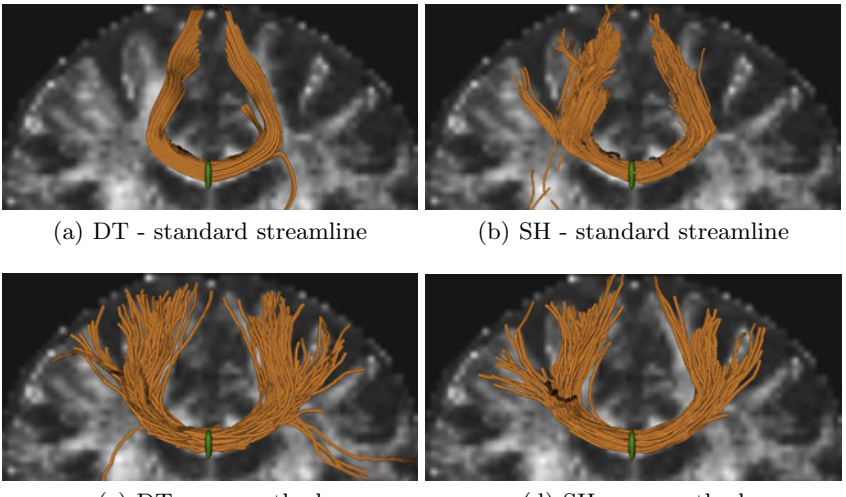
(a) Standard streamline (b) Our method

Fig. 2. Top: a slice though the synthetic fODF volume. Bottom: tractography result using standard streamline tractography [3] (a) and using our method (b). The tractography is overlaid on an image displaying generalized FA. Both tracts were seeded in the bottom left corner of the path.

MR signal was simulated over a $15 \times 30 \times 5$ voxel grid using a single tensor model for each fibre population, with eigenvalues $[1200, 100, 100] * 10^{-6}$ s/mm^2, and background tensors with eigenvalues $[1500, 1500, 1500] * 10^{-6}$ s/mm^2. The signal from both fibre populations was then added in the crossing regions, followed by the addition of Riccian noise over the entire volume to achieve a signal-to-noise ratio of 7. In a second step, an icosahedral sampling of the unit hemisphere was used to generate a set of 81 gradient directions. Together with the synthetic diffusion-weighted signal, they were incorporated in the reconstruction scheme of [13] to produce an fODF volume, illustrated in Fig. 2. We compared our particle filter tractography method with a standard streamline method [3] on this dataset, with results shown in Fig. 2. Our method was run with $K = 1000$ particles, and a neighborhood $\mathcal{N}$ size of $5 \times 5 \times 5$ voxels. Both methods were seeded with a single seed. No major difference between the methods is observed, although our method gives a somewhat smoother path.

4.2 *In vivo* Data Validation

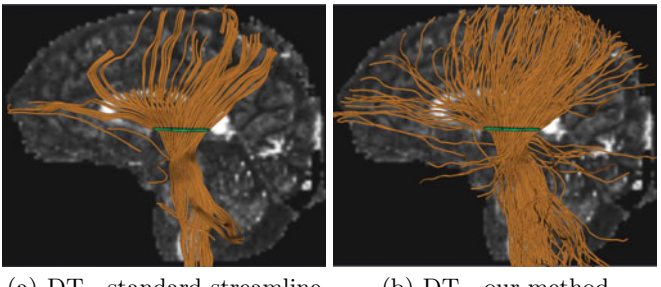
Diffusion-weighted images of a human brain were acquired on a GE Signa HDxt 3.0T scanner using an echo planar imaging sequence with a double echo option, an 8 Channel coil and ASSET with a SENSE-factor of 2. The acquisition



(a) DT - standard streamline (b) SH - standard streamline

(c) DT - our method (d) SH - our method

Fig. 3. Tractography originating from a small seed region (green) in the midsagittal slice of the corpus callosum. Top row: standard streamline tractography [3] with DT data (a) and SH data (b). Bottom row: Our particle filter tractography method applied to DT data (c) and SH data (d). The FA image is shown in the background for gross anatomic reference. Note the tracts are not exactly in the plane of the FA image and are located more anteriorly.



(a) DT - standard streamline (b) DT - our method

Fig. 4. Tractography on DT data originating from a seed region (green) in the corticospinal tract. (a): Standard streamline tractography [3]. (b): Our particle filter tractography method The FA image is shown in the background for gross anatomic reference. Note the tracts are not in the plane of the FA image and are located more laterally.

consisted in 51 directions with b = 900 s/mm^2, and 8 images with b = 0 s/mm^2, with scan parameters TR=17000 ms, TE=78 ms, FOV=24 cm, 144 × 144 encoding steps, 1.7 mm slice thickness, 85 axial slices covering the whole brain.

We demonstrate our tractography method both on DT data as well as on a spherical harmonic (SH) reconstruction of the fODF obtained with the method of [13], with order 8 spherical harmonics and a regularization parameter $\lambda = 0.006$. Tractography was initiated in a small region of interest (ROI) located in the

midsagittal slice of the corpus callosum, with 10 streamline seeds per voxel. We ran our particle filter method using $K = 1000$ particles for the DT example, and $K = 100$ particles for the SH data. A smaller number of particles for SH data is sufficient possibly because more directional information is available as compared to DT. The size of neighborhood $\mathcal{N}$ was $5 \times 5 \times 5$. Tractography was stopped when the fractional anisotropy (FA) (or the generalized FA for the case of SH data) was below a threshold, or when tractography reached the border of the brain mask. For comparison purposes, we repeated the experiments with the same seeds on the same datasets using a standard streamline tractography method [3], with the same stopping criteria. In the case of SH data, a 'branching' scheme was added to the standard method, so that when more than one peak is detected in the ODF, another streamline is seeded at that location. It can be noted from the results in Fig. 3 that our method substantially outperforms the standard method by recovering the transcallosal projections, with both types of datasets. False positive fibres obtained on the SH data by the standard method were removed using exclusion ROIs.

The methods were also compared on DT data with a seed ROI located in the cortico-spinal tract (CST) of the left hemisphere, with 2 streamline seeds per voxel and $K = 1000$ particles for our method. The results are shown in Fig. 4. Again, our method substantially outperforms standard streamline tractography. Although the *in vivo* evaluation presented here is only qualitative, it does demonstrate the recovery of tracts that are anatomically known to be present (e.g. [14]), but that the standard method fails to recover.

5 Discussion and Conclusion

This paper introduced a novel method for performing fibre tractography, formulated in a manner independent of the underlying diffusion model. The method makes use of a particle filter which estimates, at each position along the fibre, the best streamline flow model for the local fibre tract geometry. Tractography is then performed in the space of this locally estimated model, which results in the regularization of the fibre path and the ability to proceed through regions of partial volume effects, both with with diffusion tensor and high angular resolution diffusion data. This aspect of the algorithm is important, since it can result in significantly improved tractography in contexts when only low angular resolution acquisitions are available, e.g. in clinical settings. A current limitation of the method is that it propagates a single path and does not fully utilize all the information inherent in the local SF field. In future work, we will augment the algorithm with branching capabilities in locations where the estimated SF model indicates a fanning fibre configuration. The main drawback of our method with respect to traditional streamline tractography is the computational cost, proportional to the number of particles and associated mostly with computing (3). However, our results indicate that even a relatively small number of particles can produce improved tractography. Our *in vivo* results are not as smooth as those obtained with the standard method, but a more efficient implementation should allow for a larger number of particles and thus smoother tracts.

Acknowledgements. Work supported by NIH grants R01MH074794, R01MH50740, K05MH070047, P50MH080272-01, P41RR13218, U54EB005149, U41RR019703, Department of Veteran Affairs Merit Awards, VA Schizophrenia Center.

References

1. Zhang, F., Hancock, E.R., Goodlett, C., Gerig, G.: Probabilistic white matter fibre tracking using particle filtering and von Mises-Fisher sampling. Medical Image Analysis 13, 5–18 (2009)
2. Malcolm, J.G., Michailovich, O., Bouix, S., Westin, C.F., Shenton, M.E., Rathi, Y.: A filtered approach to neural tractography using the Watson directional function. Medical Image Analysis 14, 58–69 (2010)
3. Mori, S., Crain, B.J., Chacko, V.P., van Zijl, P.C.M.: Three dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Annals of Neurology 45, 265–269 (1999)
4. Lazar, M., Weinstein, D.M., Tsuruda, J.S., et al.: White matter tractography using diffusion tensor deflection. Human Brain Mapping 18(4), 306–321 (2003)
5. Parker, G., Alexander, D.: Probabilistic anatomical connectivity derived from the microscopic persistent angular structure of cerebral tissue. Phil. Trans. R. Soc. B 360, 893–902 (2005)
6. Kaden, E., Knosche, T.R., Anwander, A.: Parametric spherical deconvolution: Inferring anatomical connectivity using diffusion MR imaging. NeuroImage 37, 474–488 (2007)
7. Bjornemo, 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)
8. 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)
9. Savadjiev, P., Campbell, J.S.W., Pike, G.B., Siddiqi, K.: Streamline flows for white matter fibre pathway segmentation in diffusion MRI. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 135–143. Springer, Heidelberg (2008)
10. Barbosa, J.M., Dajczer, M., Jorge, L.P.: Minimal ruled submanifolds in spaces of constant curvature. Indiana University Mathematics Journal 33, 531–547 (1984)
11. Doucet, A., de Freitas, N., Gordon, N.: Sequential Monte Carlo methods in practice. Springer, Heidelberg (2001)
12. Arulampalam, M.S., Maskell, S., Gordon, N., Clapp, T.: A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. IEEE Trans. Signal Processing 50(2), 174–188 (2002)
13. Descoteaux, M., Angelino, E., Fitzgibbons, S., Deriche, R.: Regularized, fast and robust analytical Q-Ball imaging. Magn. Res. Medicine 58(3), 497–510 (2007)
14. Nieuwenhuys, R., Voogd, J., van Huijzen, C.: The Human Central Nervous System: A Synopsis and Atlas, 3rd edn. Springer, Heidelberg (1988)

Accurate Definition of Brain Regions Position through the Functional Landmark Approach

Bertrand Thirion[1,2], Gaël Varoquaux[1,2], and Jean-Baptiste Poline[1,2]

[1] Parietal team, INRIA Saclay-Île-de-France, Saclay, France
bertrand.thirion@inria.fr
http://parietal.saclay.inria.fr/
[2] CEA, DSV, I[2]BM, Neurospin, Saclay, France

Abstract. In many application of functional Magnetic Resonance Imaging (fMRI), including clinical or pharmacological studies, the definition of the location of the functional activity between subjects is crucial. While current acquisition and normalization procedures improve the accuracy of the functional signal localization, it is also important to ensure that functional foci detection yields accurate results, and reflects between-subject variability. Here we introduce a fast functional landmark detection procedure, that explicitly models the spatial variability of activation foci in the observed population. We compare this detection approach to standard statistical maps peak extraction procedures: we show that it yields more accurate results on simulations, and more reproducible results on a large cohort of subjects. These results demonstrate that explicit functional landmark modeling approaches are more effective than standard statistical mapping for brain functional focus detection.

1 Introduction

With the improvement of acquisition hardware, such as parallel coils and the advent of high-field MRI scanners, functional Magnetic Resonance Imaging (fMRI) provides increasingly precise information on the spatial structure of brain activity, and is expected to ultimately yield individual mappings with a resolution of 2mm. Functional activity is currently believed to be the most accessible marker to define cortical regions *in vivo* [1]. However, as far as cross-subject activity detection is concerned, the accuracy of activation position is largely determined by anatomical normalization procedures, which are also constantly improving (see [2] for a review). The precise localization of brain foci represents an important information, in particular for pharmacological MRI, where the impact of drugs in several subjects for particular brain targets is assessed [3], or in case studies where one needs to extrapolate the position of identified foci of activity to a new individual to assess lesion impact or for surgery.

However, fMRI has not been considered as a reliable marker of brain activity so far [4,3]. This is related to two problems: one is the lack of accuracy in the location of brain regions, the other one is the lack of sensitivity to detect peaks of interest, that can thus be missed when comparing functional foci across subjects.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 241–248, 2010.

In this paper, we investigate the reliability and the reproducibility of several fMRI activation detection procedures; we do not focus on the reproducibility of supra-threshold regions, but on the positions of activity peaks, for which we introduce specific metrics. In particular, we propose a fast functional landmark detection procedure, that directly aims at modeling the spatial distribution of activity peaks. This is in contrast with more standard approaches that first compute statistical maps voxel-by-voxel, then extract the peaks of these maps. This procedure, detailed in Section 2, is conceptually simpler and faster than existing procedures that perform inter-subject activation position modeling [5,6,7]. We present two validation procedures in Section 3: one based on simulations, where the true foci positions are known: we show that, when there is some jitter in the position of functional foci across subjects, the proposed functional landmark approach clearly outperforms all peak detection procedures derived from standard group statistical maps. This gain in detection reliability is characterized by a procedure designed specifically, analogous to the Receiver Operating Characteristic (ROC) curve, that shows the (sensitivity, specificity) trade-off of peak detection techniques. We proceed with an experiment based on a very large dataset, where we show that our functional landmark detection procedure yields more reproducible positions than its alternatives by using a jackknife subsampling procedure. More details and results can be found at the following address: `http://parietal.saclay.inria.fr/research/supplementary-material-miccai-2010`.

2 A Spatial Model of Brain Functional Foci

In this paper, we consider that a first-level analysis has been performed in a dataset acquired in S subjects after normalization in the MNI space, so that individual contrast, variance and statistical maps are available for certain functional contrasts (combination of condition-related effects).

2.1 Extracting Peak Positions from Standard Group Analyzes

Most activation detection procedures rely on a univariate modeling procedure that provides group-level statistical maps; these maps are then used to detect supra-threshold regions or activity peaks. The corresponding statistics can be *i)* the classical t-test, that we call random-effects statistic (*rfx*); we also consider the use of data smoothing (12mm Full Width at Half Maximum, FWHM) *srfx*; *iii)* the conjunction statistic, that assesses that the null hypothesis can be rejected in k subjects over S [8]; here we consider 2 cases, namely $k = S/2$ (half conjunction), that we call *cjh*, and *iv)* the case $k = S$ (full conjunction), called *cjf*. To make inference results comparable across statistics, the significance has been evaluated using the same sign swap procedure [9] in all cases: this procedure performs inference on the positivity of the voxel mean effect in the population, with a built-in correction for multiple comparisons.

In most fMRI studies, a few positions in the MNI space are reported as activation foci by extracting the maps local maxima. These can be ordered by decreasing importance by considering their associated statistical value.

2.2 The Functional Landmark (FL) Model

Here, we describe our new procedure to detect functional landmarks. It is related to the approach in [6], with several differences that make it simpler and faster. It consists in 3 steps: a peak extraction procedure in individual data, the specification of a spatial model at the group level, that includes a modeling of false positives; the final specification of the group-level foci of activity, with their position and their statistical significance.

High-level descriptions of individual maps. The FL detection procedure takes as input each subject t statistical map $(\phi^s)_{s=1..S}$ obtained in the cohort. For a given map ϕ we call a *terminal blob* a set a_j of connected voxels that are above a saddle point of ϕ, and that contain a unique local maximum of ϕ. We retain only the blobs that contain at least s_{min} voxels. In the sequel, we denote the extracted terminal blobs $(a_j^s)_{j=1..J(s),s=1..S}$ where s is the individual dataset under study and $j \in [1, J(s)]$ is a region index. The peak position within a_s^j is denoted t_j^s, while the average signal within the blob is denoted ϕ_j^s. Next, ϕ_j^s is converted to a probability that the blob is indeed active $p(H_1(a_j^s)|\phi_s)$, by learning the distribution of activations through a mixture model of the corresponding map ϕ_s [10]. In this work, we consider only blobs above a threshold that corresponds to a p-value $p < 0.01$ uncorrected, and use $s_{min} = 5$.

A Dirichlet Process Model for spatial data. The next step consists in selecting the blobs at similar positions across subjects. Importantly, this is carried out in a probabilistically consistent framework i.e. by refining the p-values $p(H_i(a_j^s)), i \in \{0, 1\}$ based on spatial information:

$$p(H_i(a_j^s)|t_j^s, \phi^s) = \frac{p(t_j^s|H_i(a_j^s), \phi^s)p(H_i(a_j^s)|\phi^s)}{p(t_j^s|\phi^s)}, \tag{1}$$

where $p(t_j^s|\phi^s) = \sum_{i=0}^{1} p(t_j^s|H_i(a_j^s), \phi^s)p(H_i(a_j^s)|\phi^s)$. We further assume that $p(t_j^s|H_i(a_j^s), \phi^s) = p(t_j^s|H_i(a_j^s))$. Next we specify the spatial densities under each alternative $p(t_j^s|H_i(a_j^s))$: under H_0, the selected blobs are distributed uniformly across the brain, so that $p(t_j^s|H_0(a_j^s)) = 1/|\Omega|$, where Ω is the brain volume; under H_1, the distribution $p(t_j^s|H_1(a_j^s))$ is unknown, but is expected to be clustered in some regions of the brain. To model this density, we use a Gaussian Mixture Model with an unspecified number of components implemented through a Dirichlet Process Mixture Model (DPMM) [11]:

$$\begin{aligned} \gamma &\sim DP(\theta, \mathcal{G}), \\ (\mu_j^s, \Lambda_j^s) &\sim \gamma, \forall j[1, J(s)], \forall s \in [1, S] \\ t_j^s &\sim \mathcal{N}(t_j^s; \mu_j^s, \Lambda_j^s), \forall j[1, J(s)], \forall s \in [1, S], \end{aligned} \tag{2}$$

As a base measure $\mathcal{G}$ we choose a uniform density on the compact brain volume for the mean parameters, and an Inverse Wishart distribution for the covariance

$$\mathcal{G} = p(\mu, \Lambda) = \frac{1}{|\Omega|} W^{-1}(\Lambda; \nu\sigma^2 \mathbb{I}_3, \nu), \tag{3}$$

where σ and ν are the hyper-parameters of the model. The fundamental property of Dirichlet processes is the following: if $X_J = \{(x_j), j = 1..J\}$ is sampled from a DPMM, The predictive density of a new sample is: $x|X_J, \theta, \mathcal{G} \sim \theta\mathcal{G} + \sum_{j=1}^{J} \delta_{x_j}$. Model estimation can thus be performed fairly easily using Gibbs sampling; the algorithm iteratively samples the membership z_j^s of a_j^s from

$$\begin{aligned} &p(z_j^s = k|t_j^s, z^{-s}, \theta) \propto p(t_j^s|t^{-s}, z^{-s}, z_j^s = k)p(z_j^s = k|z^{-s}, \theta), \text{ where} \\ &p(t_j^s|t^{-s}, z^{-s}, z_j^s = k) = \begin{cases} \mathcal{N}(t_j^s; \mu_k, \Lambda_k) \text{ if } n_k^{-s} > 0, \\ \frac{1}{|\Omega|} \text{ otherwise,} \end{cases} \quad \text{and} \\ &p(z_j^s = k|z^{-s}, \theta) = \begin{cases} \frac{n_k^{-s}}{\theta + N^{-s}} \text{ if } n_k^{-s} > 0, \\ \frac{\theta}{\theta + N^{-s}} \text{ otherwise,} \end{cases} \end{aligned} \tag{4}$$

where z^{-s} and t^{-s} represents the membership and position variables for the blobs in subjects other than s; n_k^{-s} is the number of instances of $z = k$ for all regions in subjects other than s and $N^{-s} = \sum_k n_k^{-s}$; $n_k^{-s} > 0$ amounts to considering that k is a previously seen component; k is unvisited otherwise. In our setting, we include an alternative case, in which the blob a_j^s is a false positive. The sampling scheme is thus:

$$p(H_0(a_j^s)|t_j^s, \phi^s) = \frac{1}{Z}\frac{1}{|\Omega|}p(H_0(a_j^s)|\phi^s), \tag{5}$$

$$p(z_j^s = k|t_j^s, z^{-s}, \theta, \phi^s) = \frac{1}{Z}\begin{cases} \frac{\theta}{\theta+N^{-s}}\frac{1}{|\Omega|}p(H_1(a_j^s)|\phi^s) \text{ if } n_k^{-s} = 0, \\ \frac{n_k^{-s}}{\theta+N^{-s}}\mathcal{N}(t_j^s; \mu_k, \Lambda_k)p(H_1(a_j^s)|\phi^s) \text{ otherwise,} \end{cases} \tag{6}$$

where the normalizing constant Z is simply obtained by summation. The parameters $(\mu_k)_{k=1..K}$ and $(\Lambda_k)_{k=1..K}$ and the number K of classes are updated at each iteration based on the current values of (z_j^s). In this work, $Q = 1000$ iterations are used, as we found that this was sufficient to yield stable estimates. We choose $\theta = 0.5$ as usually done in the literature, $\sigma = 5$mm and $\nu = 10$. These could eventually be further optimized by cross-validation.

Drawing explicit clusters from the data. Final components are estimated by clustering all the blobs that have been assigned to the same model components in at least half of the iterations, hence can be reliably thought to belong to the same group-level component. Our algorithm provides as output the average positions $(\bar{t}_k)_{k=1..K}$ of the blobs within each cluster. To assess the evidence that the resulting clusters are indeed true landmarks, we estimate a *representativity* statistic, which is simply the expectation that an active region corresponding to that cluster can be found in any subject in the group: $\eta(k) = \sum_{s=1}^{S}\left(1 - \prod_{j:z_j^s=k} p(H_0(a_j^s)|\phi_s)\right)$. It is similar to the *population prevalence* used in [7], but here the probability of each region being active is taken into account. $\eta(k)$ takes values between 0 and S. Finally, the computational cost of the whole procedure is proportional to the total number of blobs, not to the image size, which makes it much faster than all the alternatives: : by a factor of 6 with respect to [6], by a factor of (about) 400 with respect to [7].

2.3 Validation Procedures

The quality of the model is established by kernel-based statistics that measure the discrepancy between different sets of positions. Given two sets of positions $\tau = (\tau_1, .., \tau_F)$ and $t = (t_1, .., t_D)$, we can define the following asymmetric statistic, that measures how close the values in t approximate those in τ:

$$\psi(t;\tau) = \sum_{f=1}^{F} \max_{d=1..D} \exp\left(-\frac{\|\tau_f - t_d\|^2}{2\delta^2}\right) \quad (7)$$

Now, assuming that τ represents a ground truth, $\psi(t;\tau)$ can be taken as a *sensitivity* measure, while $\psi(\tau;t)$ measures the *specificity* of the detection: these quantities are continuous approximations of the number of true and false detections. As the detected peaks are a monotonous function of some threshold parameter in all the above described procedures, both quantities do monotonously increase with respect to the corresponding threshold, so that it is possible to define (1-specificity, sensitivity) plots that are analogous to ROC curves. In our experiments, we assess the sensitivity, normalized by a factor $\frac{1}{F}$, for values of the specificity below 1, so that the (specificity, sensitivity) plots remain with the $[0,1] \times [0,1]$ interval, and area under the curve (AUC) values can be used.

When dealing with real data, no ground truth is available, but we obtain sets of positions $(t^g) = (t^g_1, .., t^g_{D(g)})$ form different subgroups $g = 1..G$; then we derive a concordance index $\kappa(t) = \frac{1}{G(G-1)} \sum_{g=1}^{G} \sum_{h \neq g} \frac{1}{D(g)} \psi(t^g; t^h)$ κ is comprised between 0 and 1, a value close to 1 indicating a perfect match between subgroups, while a value close to 0 indicates a poor correspondence. We use $\delta = 10$mm in all our experiments.

3 Experiments and Results

3.1 Simulated Data

Data simulation. We have simulated a multi-subject dataset as a set of $F = 10$ distant active regions with cone-shaped activation patterns. These regions have been chosen at arbitrary positions in the mask of the standard MNI brain template. Data from 10 subjects are simulated as activation peaks corrupted with spatially correlated noise (FWHM=7mm in each direction) to mimic the spatial structure of real fMRI datasets; the spatial resolution is 3mm. The simulated activation strength is 3 times the noise standard deviation, which corresponds to realistic values. An isotropic spatial jitter of 0, 1.5, 3mm or 6mm standard deviation in each direction is added to the individual position of the foci. 100 simulations are performed for each jitter value, and the average (specificity, sensitivity) characteristics are computed across all simulations. We compare the simulated peak positions in the dataset with those obtained from standard activation voxel-level detection statistics: random effects, with or without smoothing *srfx/rfx*, half *cjh* and full *cjf* conjunctions, and functional landmarks (*fls*).

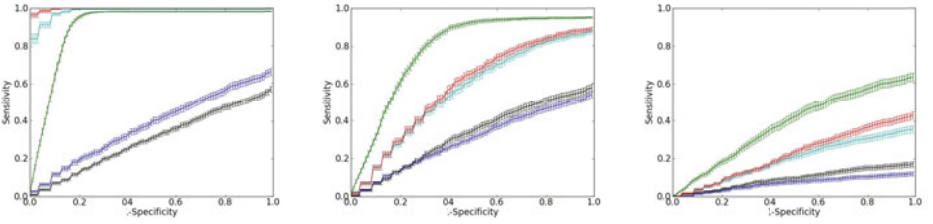


Fig. 1. (Specificity, sensitivity) curves obtained by averaging over 100 draws of simulated data, with an isotropic jitter of 0, 3 or 6mm from left to right. The blue, black, cyan, red and green characteristics correspond to *rfx, srfx, cjh, cjf* and *fls* respectively.

Table 1. Area under curves corresponding to the (specificity, sensitivity) curves shown in Fig. 1, based on 100 draws of simulated data

method	rfx	srfx	cjh	cjf	fls
AUC, jitter=0mm	0.372	0.296	0.983	**0.986**	0.898
AUC, jitter=1.5mm	0.317	0.330	0.761	0.816	**0.868**
AUC, jitter=3mm	0.297	0.323	0.560	0.577	**0.779**
AUC, jitter=6mm	0.007	0.09	0.194	0.221	**0.380**

Results: The characteristics are presented in Fig. 1 for three jitter values. The area under curve can be found in Table 1. For non-null jitter the FL approach achieves the best results, while cjh/cjf perform well only in the absence of jitter. Random effects yield poor performance, and smoothing helps only slightly in presence of jitter. A score of 0.38 (fls, jitter=6mm in each direction) means that 3.8 times more true positives than false positives are detected in average.

3.2 Experiments on Real Data

Dataset and procedure. We use here a dataset of 171 right-handed subjects, that has been described in detail in [12]. The experiment was based on an event-related fMRI paradigm that comprised ten experimental conditions. Among other tasks, the subjects had to perform left or right hand movement, which resulted in maps of the differential motor activity that we study here. FMRI data pre-processing and statistical analysis were performed using the SPM5 software (`www.fil.ucl.ac.uk`). In particular, spatial normalization was performed using default parameters (non-rigid, low frequency deformation). We use the concordance measure to compare the reproducibility of peak positions: the different position sets are obtained through jackknife subsampling, by drawing $G = 12$ random disjoint subgroups of 14 subjects from the population. Here, we use the same significance threshold, chosen as 0.05, corrected for multiple comparisons, in all procedures. Besides the statistics used in the simulation experiment, we introduced mixed-effects statistics *mfx*, as the individual variance maps were available.

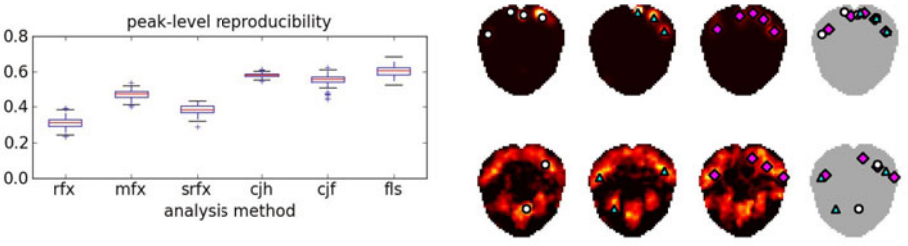


Fig. 2. (left) Reproducibility index of peak position obtained in the jackknife subsampling procedure for several group analysis techniques. (Right) 2-dimensional example on an axial slice of the scatter of peaks as observed in 3 groups, using the FL approach (top), and the *rfx* statistic (bottom): FLs are clearly more stable across groups.

Results on the real dataset. The reproducibility plots are given in Fig. 2 for all the methods tested, based on 100 random population splits into disjoint groups. We observe that the FL positions are more reproducible in average than those of the other statistics, while *rfx* and *srfx* perform worst overall. Similar results were obtained for other functional contrasts (not shown).

4 Discussion

Here we have proposed a fast procedure for functional landmark detection and characterization. The approach is conceptually simpler and faster than the alternatives proposed in [5,6,7]; on a standard PC, the procedure takes a couple of minutes. This is crucial to allow cross-validation procedures, either to assess the merit of the method in various simulation experiments, such as those presented here, or merely to optimize the parameters used in the model. Importantly also, all the parameters are continuous and can be interpreted physically, so that their impact on the results can be assessed easily, and their choice can be guided by domain knowledge. Note that all the code used here is freely available in the nipy software suite (`http://neuroimaging.scipy.org/site/index.html`).

Going back to the problem of accurate activation position detection, the FL approach outperforms voxel-based statistics approach, except in simulations with no spatial jitter. This is expected, as voxel-based statistics are straightforwardly optimal in that case, but this is not a realistic situation. The FL method is less sensitive to jitter in the cross-subject activation position than the alternative approaches. This is confirmed by the experiments on real data, that show that the FL approach generally yields more stable results than its voxel-based alternatives. The relatively poor performance of *rfx* statistic can be explained by the fact that it down-weights regions where between-subject variability is large; as these regions coincide with activated regions, this reduces the power and the reproducibility of *rfx*. Smoothing does not help a lot in that respect. Conjunction statistics (especially *cjh*) perform much better, while mixed effects

are in-between. It is important to notice that all the approaches tested here have the same control of false detections: $p < 0.05$, corrected for multiple comparisons.

On a more technical note, we can notice that the reproducibility metric κ as well as the ψ measures on simulated data provide a meaningful and sensitive comparison of methods. One possible shortcoming of κ is that it tends to provide higher values when more detections are performed, simply by chance. However, this effect can easily be controlled by comparing the actual κ values to a value obtained by permutation, thus tabulating the value of κ under the null hypothesis.

Conclusion. We have introduced a new functional landmark procedure for inter-subject detection of functional regions, that explicitly models the statistical distribution of activity peaks, and provides more reliable foci positions than standard group statistical maps. Our detection procedure is fast and its parameters are easy to calibrate; it can thus be used easily in practical settings. As future work, we plan to extend this approach to cortical maps in order to cumulate the advantage of functional landmarks and accurate brain surface modeling.

References

1. Brett, M., Johnsrude, I.S., Owen, A.M.: The problem of functional localization in the human brain. Nat. Rev. Neurosci. 3(3), 243–249 (2002)
2. Klein, A., Andersson, J., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage 46(3), 786–802 (2009)
3. Mitsis, G.D., Iannetti, G.D., Smart, T.S., Tracey, I., Wise, R.G.: Regions of interest analysis in pharmacological fMRI: how do the definition criteria influence the inferred result? Neuroimage 40(1), 121–132 (2008)
4. Duncan, K.J., Pattamadilok, C., Knierim, I., Devlin, J.T.: Consistency and variability in functional localisers. Neuroimage 46(4), 1018–1026 (2009)
5. Coulon, O., Mangin, J.F., et al.: Structural group analysis of functional activation maps. Neuroimage 11(6 pt 1), 767–782 (2000)
6. Thirion, B., Tucholka, A., et al.: High level group analysis of fMRI data based on dirichlet process mixture models. Inf. Process Med. Imaging 20, 482–494 (2007)
7. Xu, L., Johnson, T.D., Nichols, T.E., Nee, D.E.: Modeling inter-subject variability in fMRI activation location: a bayesian hierarchical spatial model. Biometrics 65(4), 1041–1051 (2009)
8. Heller, R., Golland, Y., et al.: Conjunction group analysis: an alternative to mixed/random effect analysis. Neuroimage 37(4), 1178–1185 (2007)
9. Mériaux, S., Roche, A., Dehaene-Lambertz, G., Thirion, B., Poline, J.B.: Combined permutation test and mixed-effect model for group average analysis in fMRI. Hum. Brain Mapp. 27(5), 402–410 (2006)
10. Everitt, B.S., Bullmore, E.T.: Mixture model mapping of the brain activation in functional magnetic resonance images. Hum. Brain Mapp. 7(1), 1–14 (1999)
11. Neal, R.M.: Markov chain sampling methods for dirichlet process mixture models. Journal of Computational and Graphical Statistics 9(2), 249–265 (2000)
12. Pinel, P., Thirion, B., et al.: Fast reproducible identification and large-scale databasing of individual functional cognitive networks. BMC Neurosci. 8, 91 (2007)

A Comparison of the Cingulum Tract in ALS-B Patients and Controls Using Kernel Matching

Sander van Noorden[1], Matthan Caan[1,2], Maaike van der Graaff[2], Lucas van Vliet[1], and Frans Vos[1,2]

[1] Quantitative Imaging Group, Delft University of Technology, NL
[2] Academic Medical Center, University of Amsterdam, NL
m.w.a.caan@amc.uva.nl

Abstract. Amyotrophic lateral sclerosis (ALS) is a progressive motor neuron disease with poor prognosis. Previous DW-MRI based studies in ALS on WM tracts showed a decrease of FA in tracts related to the motor system. Recent evidence suggests that extra-motor tracts are also affected by ALS. This paper aims to analyse the cingulum tracts of ALS patients and controls. To do so, we introduce kernel matching, a novel method to obtain optimal correspondence between the white matter tracts. The orientation of tract tensors in atlas space as well as the global tract shape are employed as prior information. The method proved successful to reduce the large variance of tensor shape features along the cinguli emanating from registration errors. Only after applying the proposed kernel matching method we found a significant increase in the tensor norm of both cinguli. We hypothesize that the degeneration of fibers increases tensor norm.

1 Introduction

Amyotrophic lateral sclerosis (ALS) is a progressive motor neuron disease featuring upper as well as lower motor neuron degeneration and poor prognosis. The exact cause of ALS is currently unknown.

Diffusion weighted MRI (DW-MRI) is a powerful modality to study the degeneration of white matter (WM) tracts. Previous DW-MRI based studies on ALS performed either voxel-based analysis (VBA) of WM tracts after a full-tensor non-rigid registration [1] or tract-based spatial statistics (TBSS) after a fractional anisotropy (FA)-based registration [2]. Both studies showed a decrease of FA in tracts related to the motor system such as the corticospinal tract (CST). Sage [1] reported an increase of the mean diffusivity along the CST.

Recent evidence suggests that extra-motor tracts are also effected by ALS. For instance, both [1] and [2] found a lower FA in the body of the corpus callosum (CC). Related fMRI studies showed a decrease of blood oxygen level-dependent (BOLD) contrast in the medial part of the cingulum during the execution of motor tasks [3]. So far the cingulum has not been reported to be structurally involved in ALS.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 249–256, 2010.

The analysis of atlas-tract statistics, such as in [4], relies on a good registration. Particularly, the assessment of smaller tracts is rather sensitive to registration errors, which increase the variance in the tensor shape features. These registration errors result from spatial regularization and ignore small local mismatches in order to find a transformation that is continuous at a larger scale [5]. TBSS provides a method that is less sensitive to registration errors as it projects tensor features on the closest location of the FA skeleton. While doing so, differences in orientation between the projected tensor and the target on the FA skeleton are ignored. Hence, this may result in a misregistration for closely spaced tracts such as the cingulum and the CC. Alternatively, fibers may be tracked in subject space, transformed to the atlas space followed by clustering [6]. Although more accurate, this method is less precise due to the low SNR of the subjects' datasets in which tracking is performed.

This paper aims to analyse the cingulum tracts of ALS patients and controls. Our study employs a novel kernel matching approach that is used for registration of the cingulum tracts. The kernel matching locally improves tract correspondence by employing the orientation of tract tensors in atlas space as well as the global tract shape as prior information. By doing so a projection step such as in TBSS is rendered superfluous. Statistics are computed in subject space to avoid unnecessary interpolation of the tensor data.

2 Method

2.1 Tract Segmentation in a Population-Specific Atlas

We consider tract segmentation in a population-specific tensor atlas as an initialization for our kernel matching method. This atlas contains the complete diffusion information [7] and is built for this study using a non-rigid viscous fluid registration model [5]. Fiber tractography is performed in the atlas using the FACT algorithm [8]. The resulting fiber points are clustered and a spline is fitted through the cluster center points $\mathbf{c}_{\text{atlas}}$, resulting in a smooth centerline [9]. Tracts are represented by this set of cluster centers points, the original fiber points and the cluster-membership of each fiber point to a cluster center defined in [9] as the Gaussian-weighted distance ($\sigma = 3$mm, approx. 1 voxel) . The inverse of each subject-to-atlas transformation is computed, which effectively transforms atlas tracts back to subject space.

2.2 Optimize Cluster Location in Subject Space by Kernel Matching

Kernel matching aims at improving the atlas tract to subject transformation. The atlas tract may be warped in the normal plane $\tau(\mathbf{c}_{\text{atlas}})$ to its local orientation, to achieve correspondence with the local subject tensor orientation. From the tensors, only orientation information is employed. Given a principal eigenvector $\mathbf{v_i} = (v_x, v_y, v_z)^T$, an orientation mapping can be defined by considering the outer-product $\mathbf{M_i} = \mathbf{v_i v_i^T}$. The resulting 9-tuple can be reduced to a 5-dimensional normalized vector $\mathbf{w_i}$ [10]:

$$\mathbf{w_i} = \sqrt{\frac{3}{4}}\left(v_x^2 - v_y^2, 2v_xv_y, 2v_xv_z, 2v_yv_z, \frac{1}{\sqrt{3}}\left(2v_z^2 - v_x^2 - v_y^2\right)\right). \quad (1)$$

This operation ensures that antipodal vectors are mapped to the same point in the 5-dimensional space. We define the similarity $0 \leq s \leq 1$ in orientation between two diffusion tensors as $s(\mathbf{w}_1, \mathbf{w}_2) = \frac{1}{2}((\mathbf{w}_1 \cdot \mathbf{w}_2) + 1)$. The location of an atlas cluster center point is updated by adjusting $\mathbf{q} \in \tau(\mathbf{c}_{\text{atlas}})$ such that a matching kernel η is found in a subject that maximizes kernel similarity, $\arg\max_{\mathrm{q}} \mathrm{S}(\mathbf{q})$. $S(\mathbf{q})$ is defined as:

$$S(\mathbf{q}) = \left\{\sum_{p\in\eta} s(\mathbf{w}_{\mathbf{p},\text{atlas}}, \mathbf{w}_{\mathbf{p}+\mathbf{q},\text{subject}}) \cdot s(\mathbf{w}_{\mathbf{p},\text{atlas}}, \mathbf{w}_{\mathbf{c}_{\text{atlas}}})\right\}. \quad (2)$$

In this equation, the first term decribes the similarity in orientation between atlas and subject. The second term ensures that adjacent bundles with a different orientation are ignored.

The cubic kernel η gives an equal weight to all elements. η is sized $5 \times 5 \times 5$ voxels ($1.0 \times 1.0 \times 1.1$ cm^3), which is slightly broader than the average width of the small WM tracts which we are interested in. We limit η to WM voxels that have FA > 0.2 and mode > 0 (see below) such that their principal eigenvector is well-defined.

Contrast in WM tract orientation is largest in the direction perpendicular to the tract. Furthermore, the registration error of the atlas tract is expected to be smaller than 1.5 cm. Therefore, we update $\mathbf{c}_{\text{atlas}}$ in the plane $\tau(\mathbf{c}_{\text{atlas}})$ normal to the local tract orientation. $\tau(\mathbf{c}_{\text{atlas}})$ is of small size 1.5×1.5 cm^2, discretized at a finer 0.5 voxel-size interval resulting in 17×17 points.

Interpolation of tensor values is done in a Log-Euclidean space to prevent tensor swelling effects [11]. The atlas tensor field is transformed to subject space before $S(\mathbf{q})$ is evaluated and tensors are rotated accordingly.

Cluster locations in subject space are updated to the position $\mathbf{c}_{\text{atlas}} + \mathbf{q}$ where $S(\mathbf{q})$ is maximal. The resulting centerline through these updated cluster locations is regularized by fitting a b-spline followed by projecting each cluster to the closest point on the spline. One iteration of the algorithm was performed, as more repetitions did not alter the solution. The location of fiber points is updated by interpolation of the cluster translations using the cluster-membership [9] as weights.

2.3 Tensor Shape and Tract Statistics

Conventionally, DW-MRI analyses of WM tracts rely on the FA to describe the tensor shape. Recently, it was shown that the tensor shape is entirely described by three features: FA, mode and norm, which form an orthogonal set of features [12]. FA measures the anisotropy, mode the *type* of anisotropy (ranging from -1 for planar to $+1$ for linear anisotropy) and norm the amount of diffusivity.

These features are calculated at cluster centers as a weighted sum of the feature at all fiber points employing the cluster-membership as weights. Patient

and control feature profiles are compared per cluster center using a t-test with a modest significance threshold of $p < 0.05$ (uncorrected), because of a small population size.

3 Results

3.1 Data Acquisition

Over a period of 3 years we consecutively included all 11 patients that entered four neuromuscular outpatient clinics in the Netherlands (University Medical Centres of Amsterdam, Utrecht, Rotterdam, and the Catharina Hospital in Eindhoven) and met our inclusion criteria (ALS, bulbar onset of disease) as well as 11 age matched controls. DTI data were acquired on a 3T scanner (Intera, Philips Healthcare, Best, The Netherlands). The spatial resolution was $2.0 \times 2.0 \times 2.2$ mm, per patient 64 axial slices of matrix size 128x128 were acquired for 32 gradient directions 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 [13].

3.2 Tract-Based Spatial Statistics

TBSS was performed after tensor-based registration (section 2.1). Results are shown in fig. 1. A decreased FA in the corticospinal tract (CST) and in the body of the corpus callosum (CC) is observed. The norm is increased in a few

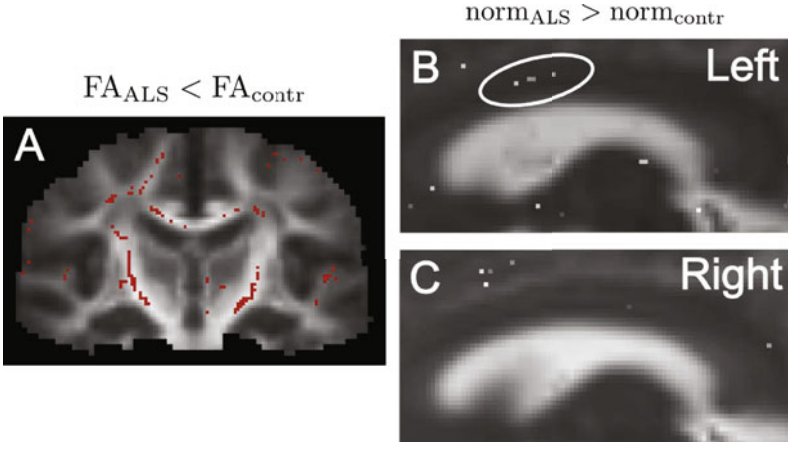


Fig. 1. TBSS results on FA (a) and norm of the left (b) and right (c) cingulum. Marked voxels show a significant decrease in FA (a) or a significant increase in norm (b,c) between ALS patients and controls according to a t-test with $p < 0.05$ (uncorrected). (a) Shows the FA is lowered for the CST and CC. (b) Shows a small effect for the norm of the left cingulum in the annotated area. (c) No significant effect is observed in the right cingulum.

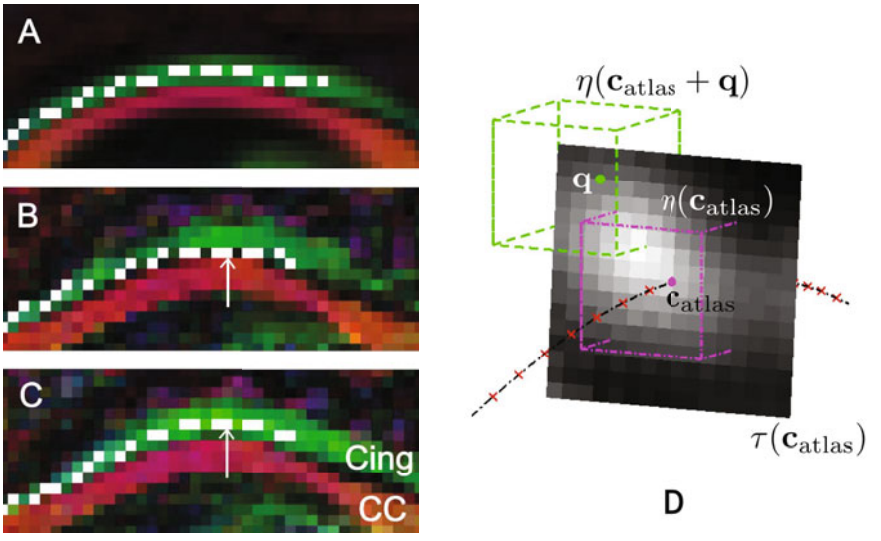


Fig. 2. Location of right cingulum cluster centers in atlas (a) and in subject before (b) and after (c) kernel matching. (d) Kernel similarity $S(\mathbf{q})$ evaluated for $\tau(\mathbf{c}_{\text{atlas}})$ around marked cluster in (b), with the kernel $\eta(\mathbf{c}_{\text{atlas}} + \mathbf{q})$ displayed at an arbitrary location.

voxels of the left cingulum, whereas no differences could be found in the right cingulum. According to the authors, TBSS is less suited for this particular area as the projection direction towards a skeleton with a tubular rather than a sheet-like topology, such as the inferior part of the cingulum, is not uniquely defined. In these areas the TBSS method employs a local search for maximum FA within a circular space in the appropriate axial slice, rather than along a single perpendicular direction [14].

3.3 Kernel Matching

Both cingulum tracts were tracked in the atlas and a centerline through its cluster center points was derived. The atlas cluster center points, depicted in fig. 2a, were transformed to subject space. Fig. 2b illustrates that in a representative subject these cluster center points did not align with the cingulum but instead were incorrectly positioned in between the cingulum and CC, due to a small registration error.

Subsequently, kernel matching was applied to obtain better correspondence between the atlas tracts and the subjects' tensor orientation field. Figure 2d shows the kernel similarity $S(\mathbf{q})$ evaluated in the normal plane $\tau(\mathbf{c}_{\text{atlas}})$ of the annotated cluster center point in fig. 2b. A high contrast in $S(\mathbf{q})$ may be observed for a small region adjacent to the initialized cluster center point $\mathbf{c}_{\text{atlas}}$. This region corresponds to the cross section of the subjects' cingulum bundle at that particular location. The initial cluster center point (purple) is clearly

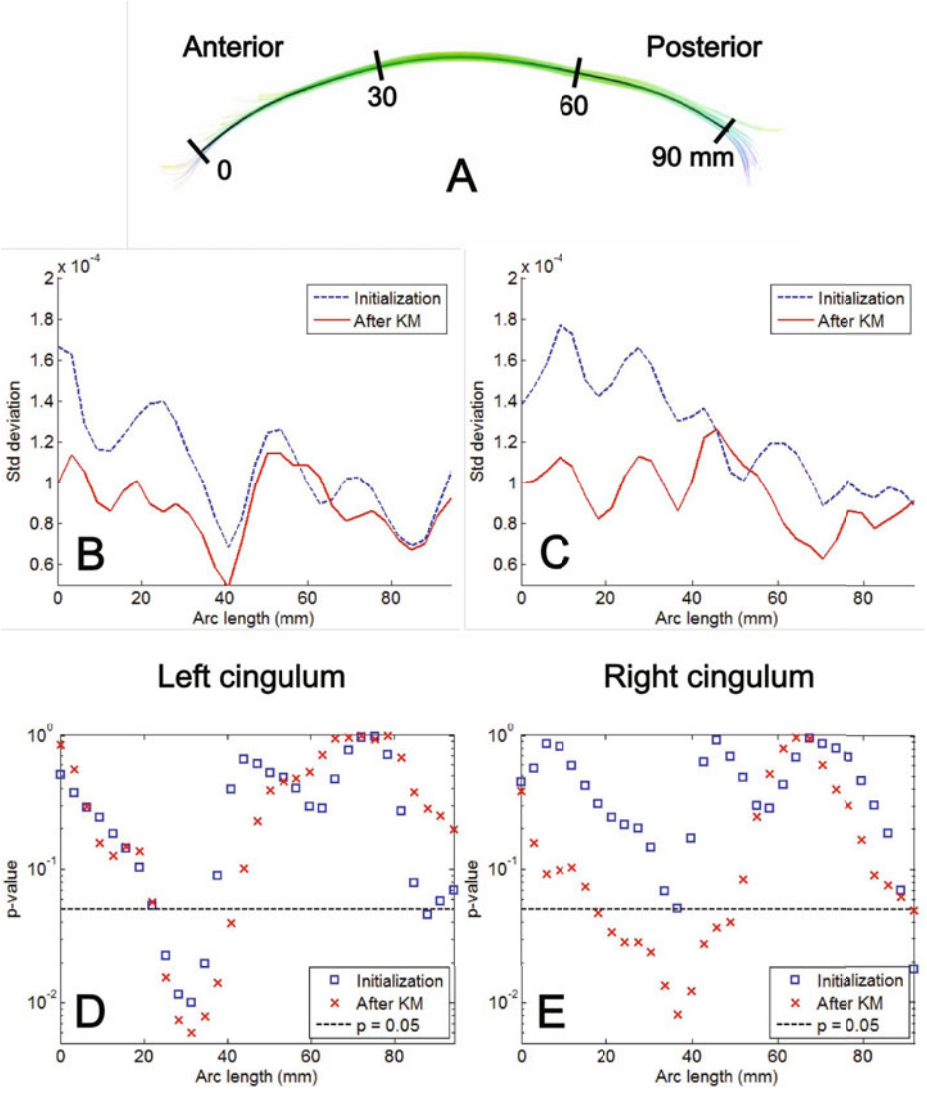


Fig. 3. (a) Defining arc length to parameterize the right cingulum in atlas space. (b,c) Standard deviation of the norm for the left (b) and right (c) cingulum in controls before and after kernel matching. Mean norm for both cinguli is approximately 1.5×10^{-3} (d,e) T-test results for the difference in norm between patients and controls evaluated at the left (d) and right (e) cingulum cluster centers. The dashed line denotes the significance threshold ($p < 0.05$).

suboptimally positioned and will be translated by $\mathbf{q}$ to the voxel with maximal $S(\mathbf{q})$. Kernel matching results in a clear improvement in correspondence between the atlas tracts and this subjects' tensor orientation field (see fig. 2c).

Tensor feature profiles of tracts evaluated after kernel matching are expected to have a lower standard deviation since these tracts are better aligned to the subject data. This reduction is demonstrated in the norm profiles of both cinguli of the control group in fig. 3b-c. The standard deviation is reduced up to 30% for clusters in the anterior and medial parts of the cingulum. Similar reductions are observed for the FA and mode profiles (data not shown).

Feature profiles that are evaluated before the application of kernel matching show an increase of tensor norm only in the left cingulum that is similar to the TBSS result, see fig. 3d. FA and mode profiles do not show an effect (data not shown). The reduced standard deviation after kernel matching enables a more accurate comparison of patient and control profiles. This results in an enlarged significance of the effects observed in the left cingulum and in addition a similar effect in the right cingulum, see fig. 3d-e. A permutation test as in [1] without correction for multiple comparisons gives exactly the same results as the paired t-test. The suprathreshold multiple comparison correction [15] indicates a significant difference of the entire right CG, but not the left.

KM was also applied to other WM tracts of interest (CST, Inferior Longitudinal Fasciculus, Arcuate Fasciculus and Fornix) resulting in similar decreases in standard deviation as in the CG tracts. No additional significant differences between controls and patients were found.

4 Discussion

Both the atlas-tract based method without kernel matching and the TBSS analysis showed a difference in norm for the left cingulum. Only after applying the proposed kernel matching method we found a significant increase in the tensor norm of the right cingulum. This effect may be attributed to a higher initial variance, reduced after kernel matching. We hypothesize that the degeneration of fibers increases tensor norm. A related increase in tensor FA is also found but not significant, which may be due to the low number of included patients.

Future work might extend the application of our method to improve the correspondence of more sheet-like tracts by applying kernel matching to the skeleton points defined in the continuous medial representation suggested by [16]. Here, the matching may be restricted to translate cluster center points along the normal line to the sheet structure, instead of the normal plane of the fiber bundle.

References

1. Sage, C.A., Van Hecke, W., Peeters, R., Sijbers, J., Robberecht, W., Parizel, P., Marchal, G., Leemans, A., Sunaert, S.: Quantitative diffusion tensor imaging in amyotrophic lateral sclerosis: Revisited. Hum. Brain Mapp. 30(11), 3657–3675 (2009)

2. Ciccarelli, O., Behrens, T.E., Johansen-Berg, H., Talbot, K., Orrell, R.W., Howard, R.S., Nunes, R.G., Miller, D.H., Matthews, P.M., Thompson, A.J., Smith, S.M.: Investigation of white matter pathology in als and pls using tract-based spatial statistics. Hum. Brain Mapp. 30(2), 615–624 (2009)
3. Lule, D., Diekmann, V., Kassubek, J., Kurt, A., Birbaumer, N., Ludolph, A.C., Kraft, E.: Cortical plasticity in amyotrophic lateral sclerosis: Motor imagery and function. Neurorehab. Neural. Re. 21(6), 518–526 (2007)
4. Goodlett, C., Fletcher, P., Gilmore, J., Gerig, G.: Group analysis of dti fiber tract statistics with application to neurodevelopment. Neuroimage 45, S133–S142 (2009)
5. Van Hecke, W., Leemans, A., D'Agostino, E., De Backer, S., Vandervliet, E., Parizel, P.M., Sijbers, J.: Nonrigid coregistration of diffusion tensor images using a viscous fluid model and mutual information. IEEE T. Med. Imaging 26(11), 1598–1612 (2007)
6. O'Donnell, L.J., Westin, C.F., Golby, A.J.: Tract-based morphometry for white matter group analysis. Neuroimage 45(3), 832–844 (2009)
7. Van Hecke, W., Sijbers, J., D'Agostino, E., Maes, F., De Backer, S., Vandervliet, E., Parizel, P.M., Leemans, A.: On the construction of an inter-subject diffusion tensor magnetic resonance atlas of the healthy human brain. Neuroimage 43(1), 69–80 (2008)
8. Jiang, H.Y., van Zijl, P.C.M., Kim, J., Pearlson, G.D., Mori, S.: Dtistudio: Resource program for diffusion tensor computation and fiber bundle tracking. Comput. Meth. Prog. Bio 81(2), 106–116 (2006)
9. Caan, M., Van Vliet, L., Majoie, C., Aukema, E., Grimbergen, K., Vos, F.: Spatial consistency in 3d tract-based clustering statistics. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 535–542. Springer, Heidelberg (2008)
10. Rieger, B., Van Vliet, L.: A systematic approach to nd orientation representation. Image Vision Comput. 22, 453–459 (2004)
11. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-euclidean metrics for fast and simple calculus on diffusion tensors. Magn. Reson. Med. 56(2), 411–421 (2006)
12. Kindlmann, G., Ennis, D.B., Whitaker, R.T., Westin, C.F.: Diffusion tensor analysis with invariant gradients and rotation tangents. IEEE T. Med. Imaging 26(11), 1483–1499 (2007)
13. Mangin, J., Poupon, C., Clark, C., Bihan, D., Bloch, I.: Eddy-current distortion correction and robust tensor estimation for mr diffusion imaging. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 186–193. Springer, Heidelberg (2001)
14. 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.J.: Tract-based spatial statistics: Voxelwise analysis of multi-subject diffusion data. Neuroimage 31(4), 1487–1505 (2006)
15. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: A primer with examples. Hum. Brain Mapp. 15(1), 1–25 (2002)
16. Yushkevich, P.A., Zhang, H., Simon, T.J., Gee, J.C.: Structure-specific statistical mapping of white matter tracts. Neuroimage 41(2), 448–461 (2008)

Optimally-Discriminative Voxel-Based Analysis

Tianhao Zhang and Christos Davatzikos

Section of Biomedical Image Analysis, Department of Radiology,
University of Pennsylvania, Philadelphia, PA 19104, USA
{Tianhao.Zhang,Christos.Davatzikos}@uphs.upenn.edu

Abstract. Gaussian smoothing of images is an important step in Voxel-based Analysis and Statistical Parametric Mapping (VBA-SPM); it accounts for registration errors and integrates imaging signals from a region around each voxel being analyzed. However, it has also become a limitation of VBA-SPM based methods, since it is often chosen empirically, non-optimally, and lacks spatial adaptivity to the shape and spatial extent of the region of interest. In this paper, we propose a new framework, named Optimally-Discriminative Voxel-Based Analysis (ODVBA), for determining the optimal spatially adaptive smoothing of images, followed by applying voxel-based group analysis. In ODVBA, Nonnegative Discriminative Projection is applied locally to get the direction that best discriminates between two groups, e.g. patients and controls; this direction is equivalent to local filtering by an optimal kernel whose coefficients define the optimally discriminative direction. By considering all the neighborhoods that contain a given voxel, we then compose this information to produce the statistic for each voxel. Permutation tests are finally used to obtain the statistical significance. The experiments on Mild Cognitive Impairment (MCI) study have shown the effectiveness of the framework.

1 Introduction

Voxel-based Analysis and Statistical Parametric Mapping (VBA-SPM) [2] [7] of imaging data have offered the potential to analyze structural and functional data in great spatial detail, without the need to define a priori regions of interests (ROIs). A fundamentally important aspect of VBA-SPM has been the spatial smoothing of images prior to analysis. Typically, Gaussian blurs of full-width-half-max (FWHM) in the range of 8-15mm are used to account for registration errors, to Gaussianize data, and to integrate imaging signals from a region, rather than from a single voxel.

The effect of this smoothing function is critical: if the kernel is too small for the task, statistical power will be lost and large numbers of false negatives will confound the analysis; if the kernel is too large, statistical power can also be lost by blurring image measurements from regions that display group differences with measurements from regions that have no group difference. In the latter case, spatial localization is also seriously compromised, as significant smoothing blurs the measurements out and often leads to false conclusions about the origin of a

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 257–265, 2010.

functional activation or of structural abnormalities. Moreover, a filter that is too large, or that is not matched with the underlying group difference, will also have reduced sensitivity in detecting group differences. As a result, Gaussian smoothing is often chosen empirically, or in an ad hoc fashion, an obvious limitation of such VBA-SPM analyses.

However, the most profound limitation of Gaussian smoothing of images is its lack of spatial adaptivity to the shape and spatial extent of the region of interest. For example, if atrophy or functional activation in the hippocampus is to be detected, Gaussian smoothing will blur volumetric or activation measurements from the hippocampus with such measurements from surrounding tissues, including the ventricles, the fusiform gyrus, and the white matter. Some earlier work in the literature [5] had shown that spatially adaptive filtering of image data can improve statistical power to detect group differences, however it didn't offer a way to determine optimal data filtering. In general, little is known about how to optimally define the shape and extent of the smoothing filter, so as to maximize the ability of VBA-SPM to detect group effects.

In this paper, we present a mathematically rigorous framework for determining the optimal spatial smoothing of medical images, prior to applying voxel-based group analysis. We consider this problem in the context of determining group differences, and we therefore restrict our experiments to voxel-wise statistical hypothesis testing. In order to determine the optimal smoothing kernel, a local discriminative analysis, restricted by appropriate nonnegativity constraints, is applied to a spatial neighborhood around each voxel, aiming to find the direction (in a space of dimensionality equal to the size of the neighborhood) that best highlights the difference between two groups in that neighborhood. Since each voxel belongs to a large number of such neighborhoods, each centered on one of its neighboring voxels, the group difference at each voxel is determined by a composition of all these optimal smoothing directions. Permutation tests are used to obtain the statistical significance of the resulting ODVBA maps.

2 The Proposed Framework

The proposed framework contains three stages: 1) Local Nonnegative Discriminative Projection, 2) Determining each voxel's statistic, and 3) Permutation tests.

2.1 Local Nonnegative Discriminative Projection

Learning Set Construction. For a given voxel x in volume X, we construct its neighborhood $\mathbb{N}$: $\|x - x_i\| < \xi$. To render subsquent processing tractable, we randomly select $k-1$ voxels $x_1, \cdots, x_{k-1}$ in this neighborhood and represent this neighborhood using a k dimensional subvolume vector: $\boldsymbol{\theta} = [x, x_1, \cdots, x_{k-1}]^T$. Provided that there are N subjects, we can obtain N subvolume vectors which form a data set: $\Theta = [\theta_1, \theta_2, \cdots, \theta_N]$ for learning. The procedure is illustrated in Fig. 1.

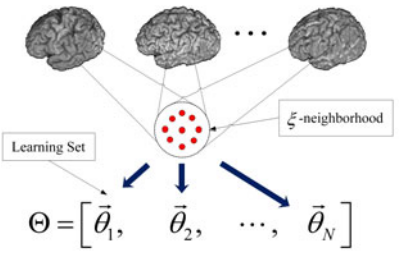


Fig. 1. Learning set construction

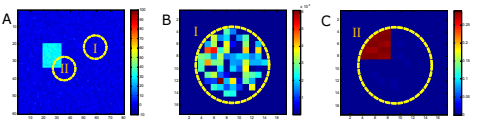


Fig. 2. Illustration of the basic idea of NDP using a toy dataset

The Basic Idea of NDP. The Nonnegative Discriminative Projection (NDP) algorithm is used to find the optimal discriminative directions which project the high-dimensional subvolume samples onto a 1-dimensional space maximizing the classification accuracy. The resultant optimally filter $\boldsymbol{w}$ is nonnegative, because of the nonnegativity constraint incorporated into the objective function. This constraint is used to help us interpret the group differences. Specifically, our goal is not simply to find an image contrast, prescribed by $\boldsymbol{w}$, which distinguishes the two groups, but also requires that this contrast tells us something about the properties of the images we are measuring, e.g. about regional volumetrics or functional activity. We therefore limit ourselves to nonnegative, albeit arbitrarily shaped, local filters, each of which prescribes a regional weighted average of the signal being measured, and therefore can be easily interpreted.

To illustrate the idea of NDP, we show its results on a toy dataset before describing the formulation. We generated two groups of images containing a square with intensity varying from one image to another: the first set of squares had intensities with mean 120.53 and standard deviation 5.79, while the second had 90.36 and 5.72, respectively. Fig. 2A shows the difference of means from the two groups. Fig. 2B shows the $\boldsymbol{w}$ obtained from the learning set constructed according to the neighborhood I; it is basically noise with very small values of $(\boldsymbol{w})_j$, indicating that no local filter can be found that distinguishes the two groups at that neighborhood. Fig. 2C shows the $\boldsymbol{w}$ obtained from the learning set corresponding to neighborhood II; the estimated $\boldsymbol{w}$ is well aligned with the underlying group difference, within which it has high values. The bottom-line here is that a properly estimated $\boldsymbol{w}$ can highlight the underlying difference.

The Formulation of NDP. Using one given learning set, we probe into neighborhood elements' contributions for discrimination of the two groups. We target

to find such a nonnegative vector $\boldsymbol{w}$: the larger the value of $(\boldsymbol{w})_j$ is, the more the corresponding element $(\boldsymbol{\theta})_j$ contributes to the discrimination. Equivalently, $(\boldsymbol{w})_j$ is the jth coefficient of the regional filter denoted by $\boldsymbol{w}$. Via $\boldsymbol{w}$, the learning set can be projected from the k-dimensional space onto the 1-dimensional space to be optimally classified, such as $\Psi = \boldsymbol{w}^T\boldsymbol{\theta}$. We expect that, in the projected space, the two classes will be separated as much as possible along $\boldsymbol{w}$, and at the same time the samples from the same class get more compact. A measure of the separation between the two classes is $\boldsymbol{w}^T S_B \boldsymbol{w}$, where $S_B = (\boldsymbol{m_1} - \boldsymbol{m_2})(\boldsymbol{m_1} - \boldsymbol{m_2})^T$; $\boldsymbol{m_i} = \frac{1}{N_i}\sum_{\boldsymbol{\theta}\in C_i}\boldsymbol{\theta}$; C_i means the ith class; N_i denotes the number of samples in C_i. And, the intraclass compactness can be described by $\boldsymbol{w}^T S_W \boldsymbol{w}$, where $S_W = \sum_{i=1}^{2}\sum_{\theta\in C_i}(\boldsymbol{\theta} - \boldsymbol{m_i})(\boldsymbol{\theta} - \boldsymbol{m_i})^T$. S_B and S_W are called the between-class scatter matrix and the within-class scatter matrix respectively, according to the classic Fisher LDA [6] in which the criterion function is based on the generalized Rayleigh quotient. Herein, S_B and S_W are considered under the formulation of quadratic programming which is amenable to the nonnegative constraint as follows:

$$\begin{aligned} &J(\boldsymbol{w}) = \min_{\boldsymbol{w}} \boldsymbol{w}^T A \boldsymbol{w} - \mu \boldsymbol{e}^T \boldsymbol{w} \\ &\text{subject to } (\boldsymbol{w})_j \geq 0, j = 1, \cdots, k, \end{aligned} \tag{1}$$

where, $A = (\gamma S_W - S_B + (|\lambda_{min}| + \tau^2)I)$; γ is the tuning parameter; $|\lambda_{min}|$ is the absolute value of the smallest eigenvalue of $\gamma S_W - S_B$; $\tau^2 << 1$ is the regularization parameter; I is the identity matrix; $\boldsymbol{e} = [1, \cdots, 1]^T$; the second term $\boldsymbol{e}^T\boldsymbol{w}$ is used to achieve $\sum_{i=1}^{k}(\boldsymbol{w})_i > 0$ which means the solutions of $(\boldsymbol{w})_i$ are not all zeros under the nonnegative constraint; μ is the balance parameter.

Theorem 1. *A is a positive definite matrix.*

Proof. If $\lambda_{min} \geq 0$, the smallest eigenvalue of A is $2\lambda_{min} + \tau^2$ which is greater than 0. If $\lambda_{min} < 0$, the smallest eigenvalue of A is just τ^2. In a words, all eigenvalues of A are greater than 0. Since S_W, S_B, and I are all symmetric matrices, A is a symmetric matrix. Thus, we complete the proof. □

Since A is positive definite, $J(\boldsymbol{w})$ is a convex function and has the unique global minimum. We solve the above optimization problem using the Nonnegative Quadratic Programming (NQP) [11]. According to [11], define the nonnegative matrices A^+ and A^- as follows: $A^+_{ij} = A_{ij}$, if $A_{ij} > 0$; otherwise, it is 0. $A^-_{ij} = |A_{ij}|$, if $A_{ij} < 0$; otherwise it is 0. So it is clear that $A = A^+ - A^-$.

Multiplicative updates rule which does not involve the learning rates, is introduced to minimize the objective function iteratively:

$$(\boldsymbol{w})_i \leftarrow \left(\frac{(\mu\boldsymbol{e})_i + \sqrt{(\mu\boldsymbol{e})_i^2 + 16(A^+\boldsymbol{w})_i(A^-\boldsymbol{w})_i}}{4(A^+\boldsymbol{w})_i}\right)(\boldsymbol{w})_i, \tag{2}$$

where $i = 1, \cdots, k$. Eq.2 means that all the elements in $\boldsymbol{w}$ are updated in parallel. Since $(A^+\boldsymbol{w})_i \geq 0$ and $(A^-\boldsymbol{w})_i \geq 0$, the updated $\boldsymbol{w}$ in Eq.2 is always nonnegative.

Theorem 2. *The function of $J(\boldsymbol{w})$ in Eq.1 decreases monotonically to the value of its global minimum under the multiplicative updates in Eq.2.*

proof. An auxiliary function ([9], pp. 659, Definition 1)([11], pp. 2013, Theorem 1) as follows is used to derive the multiple updates:

$$G(\boldsymbol{v},\boldsymbol{w}) = \sum_i \frac{(A^+\boldsymbol{w})_i}{(\boldsymbol{w})_i}(\boldsymbol{v})_i^2 - \sum_{ij} A_{ij}^-(\boldsymbol{w})_i(\boldsymbol{w})_j\left(1 + log\frac{(\boldsymbol{v})_i(\boldsymbol{v})_j}{(\boldsymbol{w})_i(\boldsymbol{w})_j}\right) - \sum_i(\mu\boldsymbol{e}^T)_i(\boldsymbol{v})_i. \tag{3}$$

According to ([9], pp. 659, Lemma 1), $J(\boldsymbol{w})$ is nonincreasing under the updates: $\boldsymbol{w} = \arg\min_{\boldsymbol{v}} G(\boldsymbol{v},\boldsymbol{w})$ and for each component in $\boldsymbol{w}$, $(\boldsymbol{w})_i = (\boldsymbol{v})_i|_{G_i'=0}$, where $G_i' = 2(A^+\boldsymbol{w})_i(\boldsymbol{v})_i/(\boldsymbol{w})_i - 2(A^-\boldsymbol{w})_i(\boldsymbol{w})_i/(\boldsymbol{v})_i - (\mu\boldsymbol{e}^T)_i$. So, we can obtain the updates described as Eq. 2. □

2.2 Determining Each Voxel'S Statistic

For all the M voxels in one volume, we have M discriminative directions, each applied to a different neighborhood, as described in Section 2.1. For a given voxel x, we obtain a list of $(\boldsymbol{w})_j$ values since x may belong to a number of neighborhoods. To quantify the group difference measured at voxel x , we use the *discrimination degree*, which relates to the effect size [3]:

$$\delta = \left(\frac{|\tilde{m}_1 - \tilde{m}_2|}{\sqrt{\sum_{i=1}^{2}\sum_{\Psi\in C_i}(\Psi - \tilde{m}_i)^2}}\sqrt{N_1+N_2-2}\right)^{\phi}, \tag{4}$$

where, $\tilde{m}_i = \frac{1}{N_i}\sum_{\Psi\in C_i}\Psi$, ϕ is the tuning parameter for reducing potential outliers in the dataset. Let $\Delta = \{\mathbb{N}|x \in \mathbb{N}\}$ denote the set of neighborhoods that a voxel x belongs to, then we define the group difference on x by summing up contributions from all neighborhoods to which it participates:

$$S_x = \sum_{\mathbb{N}\in\Delta} \delta_{\mathbb{N}}\left|(\boldsymbol{w}_{\mathbb{N}})_j\right|,\ j \in \{1,\cdots,k\}, \tag{5}$$

where, $\boldsymbol{w}_{\mathbb{N}}$ denotes the coefficients corresponding to voxels in $\mathbb{N}$, $(\boldsymbol{w}_{\mathbb{N}})_j$ denotes that x is the jth element in $\mathbb{N}$, and $\delta_{\mathbb{N}}$ which acts as the weight for $\boldsymbol{w}_{\mathbb{N}}$ denotes the *discrimination degree* achieved in neighborhood $\mathbb{N}$ and is defined in Eq. 4. S_x will serve as the statistic reflecting group differences on the voxel x, and will be used next to determine statistical significance. Higher values of S_x reflect stronger group differences.

2.3 Permutation Tests

Assume the null hypothesis that no difference between the two groups, the statistical significance can be assessed by comparison with the distribution of values

obtained when the labels are randomly permuted [10]. In particular, we randomly assign the subjects into two groups, and then implement Section 2.1- Section 2.2 to calculate the statistic for each voxel. The above relabeling is repeated N_P times. For one given voxel, let S_0 denote the statistic value obtained under the initial class labels, and $S_i, i = 1, \cdots, N_P$ denotes the ones obtained by relabeling. The P value for the given voxel is calculated according to:

$$P = \sum_{i=1}^{N_P} \left[u(S_i - S_0)\right] / N_P \tag{6}$$

where, $u(t) = 1$, if $t \geq 0$; otherwise it is 0.

3 Results

In this section, we carry out the study of determining the extent of atrophy in Mild Cognitive Impairment (MCI) subjects to evaluate ODVBA compared with the original SPM [2] and the nonparametric permutation based SPM (SnPM) [10]. The data was obtained from ADNI [1], which has recruited approximately 800 adults including 200 normal controls, 400 individuals with MCI and 200 Alzheimer's disease (AD) patients. The images were acquired and processed according to a number of steps detailed under the ADNI website [1]. We randomly selected 100 subjects with MCI from the ADNI cohort. 50 of these subjects that had undergone conversion to AD were referred to MCI-C. The remaining 50 non-converters were referred to MCI-NC. Images were preprocessed according to the following steps. 1) Alignment to the ACPC plane; 2) Removal of extracranial material; 3) Tissue segmentation into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF); 4) High-dimensional image warping to a standardized coordinate system; 5) Formation of tissue density maps typically used in the modulated SPM analysis [4]. We used GM for evaluation purposes.

Both SnPM and ODVBA are implemented with 2000 permutations. Fig. 3 shows some selected sections from the results (with P value < 0.001 thresthold) of

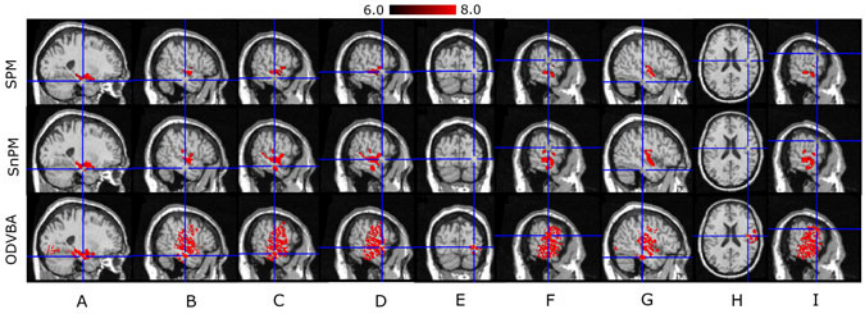


Fig. 3. Representative sections with significant regions. The scale indicates the $-log(P)$ values.

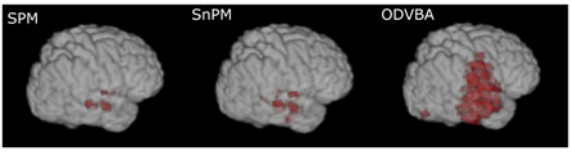


Fig. 4. Surface renderings of regions

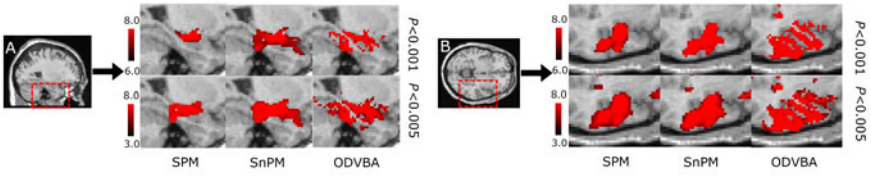


Fig. 5. Representative magnified regions with two levels of P value

SPM, SnPM, and ODVBA, respectively. We can see that the results of ODVBA reflect significant GM loss in MCI-C compared with MCI-NC mainly in Hippocampus (Fig. 3A), Inferior Temporal Gyrus (Fig. 3B), Middle Temporal Gyrus (Fig. 3C), Superior Temporal Gyrus (Fig. 3D), Occipital Lobe (Fig. 3E), Insular Cortex (Fig. 3F), Fusiform Gyrus (Fig. 3G), Parietal Lobe (Fig. 3H), and Inferior Frontal Gyrus (Fig. 3I). In addition, some significant regions detected by ODVBA are either totally or partially obscured in the results of SPM and SnPM. Surface renderings in Fig. 4 also demonstrate the results of the three methods. As we can see, ODVBA reveals a substantially larger and more detailed area of atrophy in Temporal Lobe, Parietal Lobe, and Frontal Lobe than SPM and SnPM, which barely detected the damage. Moreover, these are all regions that are generally known from histopathology studies to be affected in AD.

Fig. 5 shows two representative magnified regions that were found by the three methods. Among them, Fig. 5A shows the region near the Hippocampus and Fig. 5B shows the region around the Temporal Lobe. The results are with P value < 0.005 threshold and P value < 0.001 threshold respectively. For the Fig. 5A, we can see that SPM and SnPM blurred the regions of the Hippocampus

Table 1. Statistics of Clusters

Cluster	P-value<0.005						P-value<0.001					
	SPM		SnPM		ODVBA		SPM		SnPM		ODVBA	
	size	t	size	t	size	t	size	t	size	t	size	t
#1	10657	7.86	13463	7.65	13816	**8.47**	2430	8.12	2430	7.35	9568	**8.70**
#2	398	4.96	737	4.79	2341	**8.18**	1424	6.81	1424	6.34	855	**7.09**
#3	325	5.33	359	5.43	1657	**6.93**	1365	5.27	1365	6.99	575	**7.28**

and the Fusiform Gyrus. In contrast, a clear separation of the two regions can be found in the results of ODVBA. Moreover, the results of SPM with P value < 0.001 detected no significant atrophy located in the Fusiform Gyrus in that section. For the Fig. 5B, SPM and SnPM blurred the different gyri and sulci in the region of the Temporal Lobe; however, ODVBA delineates a more precise area of significant atrophy in that region.

In Table 1, we list the t value on the three biggest clusters with P values < 0.005 and < 0.001, respectively. The t value is calculated based on the cluster means on the tissue density maps of the two groups' samples. We can see that the t values with clusters of ODVBA are higher than those of SPM and SnPM; that is, the regions found by ODVBA display a greater difference between the two groups than SPM and SnPM.

4 Summary

We have introduced a new framework of voxel-based analysis, aiming to detect differences associated with brain abnormalities existing in two groups. The main premise of this approach is that the optimal shape and size of the spatial filter to be applied to the data prior to statistical analysis is not known in advance, but must be estimated from the data. Moreover, this spatial filtering is not fixed throughout the image, as customary in the literature, but is spatially adaptive, depending on the local anatomy and abnormality (which is unknown in advance, as well). We presented a nonnegative discriminative direction method, which determines the filter that best distinguishes the two groups being compared. This approach was evaluated in the study of contrasting MCI-C versus MCI-NC, and revealed substantially more extensive, and more significant GM atrophy in regions known to be affected by AD, whereas SPM and SnPM produced inferior results.

References

1. Alzheimer's Disease Neuroimaging Initiative, http://www.loni.ucla.edu/ADNI
2. Ashburner, J., Friston, K.J.: Voxel-based morphometry-the methods. Neuroimage 11(6), 805–821 (2000)
3. Cohen, J.: Statistical Power Analysis for the Behavioral Sciences, 2nd edn. Lawrence Erlbaum Associates, Mahwah (1988)
4. Davatzikos, C., Genc, A., Xu, D., Resnick, S.M.: Voxel-based morphometry using the RAVENS maps: Methods and validation using simulated longitudinal atrophy. NeuroImage 14(6), 1361–1369 (2001)
5. Davatzikos, C., Li, H.H., Herskovits, E., Resnick, S.M.: Accuracy and sensitivity of detection of activation foci in the brain via statistical parametric mapping: a study using a PET simulator. NeuroImage 13(1), 176–184 (2001)
6. Duda, R., Hart, P., Stork, D.: Pattern Classification, 2nd edn. Wiley, Chichester (2000)

7. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J.: Statistical parametric maps in functional imaging: a general linear approach. Hum. Brain Map. 2, 189–210 (1995)
8. Genovese, C.R., Lazar, N.A., Nichols, T.: Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870–878 (2002)
9. Lee, D.D., Seung, H.S.: Algorithms for non-negative matrix factorization. In: NIPS, vol. 13, pp. 556–562 (2001)
10. Nichols, T.E., Holmes, A.P.: Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Map. 15(1), 1–25 (2002)
11. Sha, F., Lin, Y., Saul, L.K., Lee, D.D.: Multiplicative updates for nonnegative quadratic programming. Neural. Comp. 19(8), 2004–2031 (2007)

Hippocampal Shape Classification Using Redundancy Constrained Feature Selection

Luping Zhou[1], Lei Wang[1], Chunhua Shen[2], and Nick Barnes[2]

[1] School of Engineering, The Australian National University
[2] Embedded Systems Theme, National ICT, Australia*

Abstract. Landmark-based 3D hippocampal shape classification involves high-dimensional descriptor space, many noisy and redundant features, and a very small number of training samples. Feature selection becomes critical in this situation, because it not only improves classification performance, but also identifies the regions that contribute more to shape discrimination. This work identifies the drawbacks of SVM-RFE, and proposes a novel class-separability-based feature selection approach to overcome them. We formulate feature selection as a constrained integer optimization and develop a new algorithm to efficiently and optimally solve this problem. Theoretical analysis and experimental study on both synthetic data and real hippocampus data demonstrate its superior performance over the prevailing SVM-RFE. Our work provides a new efficient feature selection tool for hippocampal shape classification.

1 Introduction

Identifying the morphological differences between anatomical shapes related to disorders is important for medical image analysis. However, this is very difficult because the data are often high-dimensional but training samples are scarce. For hippocampal shapes, it is common for the SPHARM-PDM, which represents shapes by corresponded landmarks from parameterized surfaces, to represent a hippocampus with more than 1,000 landmarks. Stacking their coordinates leads to a high-dimensional feature vector. However, the number of training data is commonly around 30-50 only. Even for the advanced classifiers such as the Support Vector Machines (SVMs), the presence of many irrelevant and noisy features can significantly deteriorate learning performance. Feature subset selection becomes a critical step in this situation.

Feature selection has been widely used in medical applications, for example, the well-known SVM-RFE (recursive feature elimination) method [1]. Despite its popularity in feature selection, SVM-RFE has three drawbacks: i) Because SVM maximizes the minimum margin between two groups, SVM-RFE is not robust against noisy data even with soft-margin SVM; ii) SVM-RFE cannot effectively

* National ICT Australia is funded by the Australian Government's Backing Australia's Ability initiative, in part through the Australia Research Council. The authors thank the OASIS team and NICTA AASEDP project for providing the data.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 266–273, 2010.

avoid selecting highly correlated discriminative features; and iii) SVM-RFE cannot flexibly deal with group-based feature selection. In landmark-based 3D representation of hippocampus, due to its continuous and overall smooth surface the change within a small area is not drastic. As a result, the coordinates of the landmarks in the area are often strongly correlated. The existence of such feature redundancy causes problems for the k-best feature selection. In the extreme case, if the most discriminative feature is duplicated several times, all of them will be selected and consequently those less discriminative but complementary features may be missed. This could significantly degrade the classification performance. Moreover, to benefit the explanation of the difference between hippocampal groups, the selection of landmarks is needed, that is, to select x, y, z coordinates (the features in the shape descriptor) of the same landmark simultaneously. Such a task may be cumbersome for SVM-RFE that uses the backward sequential selection. Additional criteria need to be imposed to combine the selection of individual coordinates, which might not be a natural extension.

In this paper we propose a new approach to select discriminative features in the hippocampal shape study. To address noisy features, we use the trace-based class separability measure as the feature selection criterion. This criterion has been shown to be robust to the small sample problem and noisy features [2]. However, this criterion cannot identify redundant features either. To overcome this problem, we propose a new redundancy-constrained feature selection (RCFS). The basic idea is to formulate the feature selection problem as a 0-1 linear fractional programming problem and impose extra constraints to avoid selecting redundant features. To achieve efficient feature selection, we study the constraints that maintain the global solvability through the *totally unimodular* (TUM) condition in integer programming, and demonstrate that hierarchically clustering features can generate qualified redundancy constraints. In addition, due to its flexibility of adding linear constraints, RCFS can be easily extended to select the landmark points. Experiments show that the proposed RCFS method significantly outperforms SVM-RFE on the hippocampus data due to its more robust selection criterion, the capability in identifying and removing redundant features, and the flexible extension for landmark selection.

2 Redundancy-Constrained Feature Selection (RCFS)

Let $(\boldsymbol{x}, y) \in (\mathbb{R}^n \times \mathcal{Y})$ be a training sample, where $\mathcal{Y} = \{1, 2, \cdots, s\}$ is the label set. Let l_i be the number of samples in class i, $\boldsymbol{m}_i$ the mean of class i and $\boldsymbol{m}$ the mean of all classes. The within-class, between-class and total scatter matrices are defined as

$$\mathbf{S}_W = \sum_{i=1}^{s} \sum_{j=1}^{l_i} (\boldsymbol{x}_{ij} - \boldsymbol{m}_i)(\boldsymbol{x}_{ij} - \boldsymbol{m}_i)^\top, \quad \mathbf{S}_B = \sum_{i=1}^{s} l_i (\boldsymbol{m}_i - \boldsymbol{m})(\boldsymbol{m}_i - \boldsymbol{m})^\top,$$
$$\mathbf{S}_T = \mathbf{S}_W + \mathbf{S}_B = \sum_{i=1}^{s} \sum_{j=1}^{l_i} (\boldsymbol{x}_{ij} - \boldsymbol{m})(\boldsymbol{x}_{ij} - \boldsymbol{m})^\top.$$

When feature dimensionality is much larger than the number of training samples, which is the case of hippocampal shape classification, the scatter matrices are rank-deficient and the determinants become zero. Hence, the trace-based form, $\mathrm{tr}(\mathbf{S}_B)/\mathrm{tr}(\mathbf{S}_T)$, is used in this paper. It is not difficult to show that

$$\mathrm{tr}(\mathbf{S}_B) = \sum_{i=1}^{s} l_i (\boldsymbol{m}_i - \boldsymbol{m})^\top (\boldsymbol{m}_i - \boldsymbol{m}) = \sum_{t=1}^{n} \left(\sum_{i=1}^{s} l_i (m_{it} - m_t)^2 \right) \triangleq \sum_{t=1}^{n} f_t \quad (1)$$

where m_{it} and m_t are the t-th feature of $\boldsymbol{m}_i$ and $\boldsymbol{m}$, respectively. Similarly,

$$\mathrm{tr}(\mathbf{S}_T) = \sum_{t=1}^{n} \left(\sum_{i=1}^{s} \sum_{j=1}^{l_i} (x_{ijt} - m_t)^2 \right) \triangleq \sum_{t=1}^{n} g_t \quad (2)$$

where x_{ijt} is the t-th feature of $\boldsymbol{x}_{ij}$. We have proved in [3] that *if the most discriminative feature t, which has the maximal f_t/g_t, is duplicated k times, feature selection by maximizing* $\mathrm{tr}(\mathbf{S}_B)/\mathrm{tr}(\mathbf{S}_T)$ *will repetitively select it k times.* Similar results exist for sufficiently correlated features.

Basic Problem. To prevent selecting discriminative but mutually redundant features, we propose the redundancy-constrained feature selection (RCFS). Let $\boldsymbol{\omega} \in \{0,1\}^n$ be an n-dimensional binary selector ("1" for being select and "0" for not). Selecting k features can be expressed as finding the optimal $\boldsymbol{\omega}$,

$$\boldsymbol{\omega}^\star = \arg\max_{\boldsymbol{\omega}} \frac{f_1\omega_1 + \cdots + f_n\omega_n}{g_1\omega_1 + \cdots + g_n\omega_n} = \arg\max_{\boldsymbol{\omega}} \frac{\mathbf{f}^\top \boldsymbol{\omega}}{\mathbf{g}^\top \boldsymbol{\omega}} \quad (3)$$

$$\text{subject to } \boldsymbol{\omega} \in \{0,1\}^n, \ \ \boldsymbol{\omega}^\top \mathbf{1} = k, \text{ and } \ \boldsymbol{\omega} \in \boldsymbol{\Omega}.$$

$\boldsymbol{\Omega}$ contains the constraints used to avoid selecting redundant features. With the Dinkelbach's algorithm [4], solving Eq.(3) iteratively solve a subproblem,

$$\begin{aligned} z(\lambda) \triangleq \quad & \max_{\boldsymbol{\omega}} \ (\mathbf{f}^\top \boldsymbol{\omega} - \lambda \mathbf{g}^\top \boldsymbol{\omega}) \\ \text{subject to } \ & \boldsymbol{\omega} \in \{0,1\}^n, \ \boldsymbol{\omega}^\top \mathbf{1} = k, \ \boldsymbol{\omega} \in \boldsymbol{\Omega}. \end{aligned} \quad (4)$$

When $z(\lambda) = 0$, the optimal solution of (4) will be the optimal solution of (3).

Global Solvability. When $\boldsymbol{\omega} \in \{0,1\}^n$, adding $\boldsymbol{\Omega}$ could make Eq.(4) very difficult to solve, even if $\boldsymbol{\Omega}$ only contains linear constraints and (4) becomes an integer linear program (ILP). ILP is much more difficult than LP, and there are no general polynomial-time algorithms. Nevertheless, if satisfying the *totally unimodular* (TUM) condition [5] , an ILP problem will reduce to an LP problem which can be easily solved. Relaxing $\boldsymbol{\omega} \in \{0,1\}^n$ to $[0,1]^n$, Eq.(4) becomes an LP problem with the feasible region defined as

$$R(\boldsymbol{\omega}) = \{\boldsymbol{\omega} : \mathbf{A}\boldsymbol{\omega} \leq \mathbf{b}, \boldsymbol{\omega} \geq 0\}. \quad (5)$$

Geometrically, $R(\boldsymbol{\omega})$ is a polyhedron. According to [5], for each integral vector $\mathbf{b}$, $R(\boldsymbol{\omega})$ is an *integral* polyhedron if and only if the matrix $\mathbf{A}$ is TUM. Because the optimal solution of an LP problem is always at one of the vertices of the polyhedron, the optima of the ILP and LP problems coincide with each other. Hence, to efficiently solve Eq.(4), $\mathbf{A}$ in (5) has to be TUM. A TUM matrix is a matrix with the determinants of all of its square submatrices being $+1$, -1, or 0. It has the following properties. (**P1**): TUM is preserved when permuting rows or columns or taking transpose; (**P2**): TUM is preserved when multiplying

a row or column by -1 or repeating a row or column; (**P3**): If $\mathbf{A}$ is TUM, $[\mathbf{A}\ \mathbf{I}]$ is TUM, where $\mathbf{I}$ is an identity matrix.

Although it is restrictive for $\mathbf{A}$ to be TUM, we show that the constraints obtained by feature clustering gives a qualified $\mathbf{A}$. Let $x_1, x_2, \cdots, x_n$ be the n features of $\boldsymbol{x}$. We define $d(x_i, x_j)$ as the "distance" between x_i and x_j that reflects their independence or complementary. It can be correlation coefficients, mutual information, or any criterion on feature redundancy. We define $d(x_i, x_j) = 1 - |\rho(x_i, x_j)|$, where ρ is Pearson correlation. Let $\mathcal{C}_1, \mathcal{C}_2, \cdots, \mathcal{C}_m$ be m clusters, forming a mutually exclusive and complete partition of the n features,

$$\{x_1, x_2, \cdots, x_n\} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots \cup \mathcal{C}_m \quad \text{and} \quad \mathcal{C}_i \cap \mathcal{C}_j = \emptyset, \quad 1 \le i < j \le m. \tag{6}$$

We enforce that at most p_i ($p_i \ge 1$) features can be selected from $\mathcal{C}_i$,

$$\sum_{x_j \in \mathcal{C}_i} \omega_j \le p_i, \quad \forall i = 1, 2, \cdots, m. \tag{7}$$

Let $(x_{r_1}, \cdots, x_{r_n})$ be a rearrangement of $(x_1, \cdots, x_n)$ according to their appearing in $\mathcal{C}_1, \cdots, \mathcal{C}_m$ and this applies to $\boldsymbol{\omega}$ too. Let $\mathbf{I}_{n\times n}$ be an identity matrix and $\mathbf{1}_{1\times c_i}$ be a row vector of 1's. $\mathbf{A}\boldsymbol{\omega} \le \mathbf{b}$ in (5) can be explicitly written as

$$\begin{pmatrix} & \mathbf{1}_{1\times n} & & \\ & -\mathbf{1}_{1\times n} & & \\ -- & -- & -- & -- \\ \mathbf{1}_{1\times c_1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{1\times c_2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}_{1\times c_m} \\ -- & -- & -- & -- \\ & \mathbf{I}_{n\times n} & & \end{pmatrix} \begin{pmatrix} \omega_{r1} \\ \omega_{r2} \\ \vdots \\ \vdots \\ \vdots \\ \omega_{rn} \end{pmatrix} \le \begin{pmatrix} k \\ -k \\ -- \\ p_1 \\ \vdots \\ p_m \\ -- \\ \mathbf{1}_{n\times 1} \end{pmatrix}. \tag{8}$$

The middle part of $\mathbf{A}$ is an *interval* matrix. It contains "0" and "1" only and has consecutive 1's in each row. Each interval matrix is TUM [5]. It can be proved that the whole $\mathbf{A}$ is TUM by using (**P1**), (**P2**) and (**P3**). Thus, the subproblem in Eq.(4) can be efficiently solved thanks to the equivalence of ILP and LP.

Constraints Generation. The above method has many algorithmic parameters, including m and $p_1, \cdots, p_m$. Optimally setting them is impractical. We propose *agglomerative hierarchical clustering* to handle it. Starting with the n features, two features (or subclusters) are merged at each level until only k clusters are left, giving a hierarchy of $n - k + 1$ levels. Then, the constrained feature selection is applied to *each* level of this hierarchy with all p_i in (8) being 1. Multi-fold cross-validation is used to identify the best selection from different levels. In doing so, i) we do not need to preset m. Instead, features are clustered at different degrees of redundancy in this hierarchy; ii) we only need to set $p_i = 1$. Because one cluster at a given level is formed by multiple clusters at preceding levels, the case of $p_i > 1$ can be implicitly approximated by a group of $p_j = 1$ in preceding levels; iii) the matrix $\mathbf{I}$ in $\mathbf{A}$ can be ignored; iv) this will not significantly slow down feature selection because only LP problems are solved and the Dinkelbach's algorithm usually terminates in a few iterations.

Table 1. Proposed redundancy-constrained feature selection (RCFS)

Input: l training samples $\{(\boldsymbol{x}_i, y_i)\}_{i=1}^{l}$ and the value of k,
Output: optimal binary selector $\boldsymbol{\omega}$ and corresponding k selected features.
Initialization:
hierarchically cluster n features (or 3D points) with correlation coefficient ρ,
establish linear constraints $\boldsymbol{\Omega}$ accordingly,
compute g_i and f_i $(i = 1, 2, \cdots, n)$ for each feature,
initialize k components of $\boldsymbol{\omega}$ as "1" and the remaining as "0",
Feature selection on each level with the Dinkelbach's algorithm:
(1) **Set** $\lambda = \mathbf{f}^\top \boldsymbol{\omega} / \mathbf{g}^\top \boldsymbol{\omega}$,
(2) **Solve** the maximization problem in Eq. (4)
(3) If $\mathbf{f}^\top \boldsymbol{\omega} - \lambda \mathbf{g}^\top \boldsymbol{\omega} < \xi$ (e.g., 10^{-4}), $\boldsymbol{\omega}$ is optimal. Otherwise, go to (1).
Cross-validation is used to identify the best selection from different levels

Landmark Selection. 3D landmark selection for hippocampal shapes is very useful for medical diagnosis and clinical interpretation. The SPHARM-PDM representation of hippocampal surfaces stacks the x, y, z coordinate values of all landmarks as a long vector. A straightforward feature selection chooses the individual coordinates instead of a 3D point as a whole. It is highly likely that, for a point, one of its three coordinates is selected but the other two are not, bringing difficulty in interpreting the selection result. A landmark-based selection is needed, in which the three coordinates of each point are selected (or not selected) together. Our proposed RCFS can handle this case effortlessly by assigning the same ω_i to the three coordinates. It can be shown that the matrix $\mathbf{A}$ is still TUM in this case. In contrast, SVM-RFE, as a backward sequential selection, cannot handle landmark selection naturally. It needs to incorporate additional criteria to evaluate the importance of a landmark as a whole. This is not as seamless as our RCFS.

3 Experiments

Synthetic data. A synthetic data set is used to illustrate the efficacy of RCFS on redundancy removal. Only 2 (x_1 and x_2) out of 52 features are statistically relevant to class labels, whereas the others are noises. x_2 is more discriminative than x_1. Two classes are sampled from $\mathcal{N}((2, 0.25)^\top, \boldsymbol{\Sigma})$ and $\mathcal{N}((2.5, 3)^\top, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (.24\ .38;\ .38\ .81)$. x_1 and x_2 are duplicated with random noise respectively. Assuming that $k = 2$ is known, we test RCFS, SVM-RFE, and the non-constrained feature selection (NCFS) on 30 training and test groups (100 vs. 500 samples). It is found that RCFS successfully selects (x_1, x_2) on 28 groups. In contrast, SVM-RFE only succeeds on 2 groups and keeps selecting x_2 and its duplicate on other groups. NCFS never succeeds and always selects x_2 and its duplicate. With all 54 features, a linear SVM obtains the test error rate $8.01 \pm 2.18\%$. With the 2 features selected by RCFS, SVM-RFE and NCFS, a linear SVM obtains $1.47 \pm 1.48\%$, $5.22 \pm 1.39\%$ and $5.45 \pm 0.82\%$, respectively. As shown, RCFS outperforms both SVM-RFE and NCFS.

Table 2. Comparison of classification *with* and *without* RCFS feature selection

k	RCFS wins (groups)	RCFS loses (groups)	Mean (test errors %)		p-value (one tailed)
			RCFS	Use all features	(paired t-test)
2500	18	8	38.11	39.31	0.1421
2000	14	9	38.93	39.31	0.3533
1500	17	12	36.98	39.31	**0.0189**
1000	18	8	**36.42**	39.31	**0.0022**
500	18	11	37.23	39.31	0.0687
$\sum$	84	48	-		

Hippocampi in OASIS. We apply our RCFS method to improving the discrimination of hippocampal shapes between AD and the normal control. Subjects aged from 60 to 79 in the OASIS data set (http://www.oasis-brains.org/) are used. We categorize subjects with a non-zero CDR rating into the AD group and the rest into normal control. There are 103 samples for the left and right hippocampi respectively. Each shape is represented by x, y, z coordinates of 1002 landmarks (3006 features in total) obtained from SPHARM-PDM representation with degree 15. Experimental results are reported only for the left hippocampi[1]. Samples are randomly partitioned into 30 training and test (50 vs. 53 samples) groups. With all 3006 features used, a linear SVM attains an average error rate of 39.31%. Due to the complexity of data and the scarcity of training samples, the test error rates of different groups vary significantly: from 26% to 55%. This inter-group variation may hide the true difference between different methods. To give a fair and accurate evaluation, we report the number of groups on which RCFS wins or loses in addition to the average test error rates. More importantly, we conduct a paired t-test to test the statistical difference between two methods. By pairing the test error rates, each time the two methods are compared on the same data set, which mitigates the influence of the inter-group variation.

The paired t-test is first used to detect the statistical difference between the test error rates from a linear SVM using the RCFS-selected features and all 3006 features, respectively. As shown in Table 2, significant difference is detected at the level of 0.05 on 30 test groups for $k = 1000$ and 1500, at the level of 0.1 for $k = 500$. This verifies that when a suitable number of features are selected, employing RCFS can significantly improve classification accuracy. For example, using only 1/3 of the original features can reduce the average test error rate from 39.31% to 36.42%.

The paired t-test is then used to detect the statistical difference between the test error rates from a linear SVM using the features selected by RCFS and SVM-RFE, respectively. As shown in Table 3, RCFS wins much more often than SVM-RFE does. The lowest average error rate 36.42% is achieved by RCFS when

[1] Results for the right hippocampi (with higher classification accuracy than the left) are omitted here due to the limit of pages. The hypothesis test shows that the performance of RCFS statistically equals that of SVM-RFE on the right hippocampi.

Table 3. Performance comparison of RCFS and SVM-RFE

k	RCFS wins (groups)	RCFS loses (groups)	Mean (test errors %)		p-value (one tailed)
			RCFS	SVM-RFE	(paired t-test)
2500	14	8	38.11	39.06	0.1218
2000	18	9	38.93	39.31	0.3710
1500	17	8	36.98	38.99	**0.0276**
1000	18	7	**36.42**	39.75	**0.0001**
500	15	9	37.23	38.87	0.0958
$\sum$	82	41	-		

$k = 1000$, as shown in bold. More importantly, the paired t-test indicates that, RCFS and SVM-RFE are significantly different at the level of 0.001 on the 30 test groups when $k = 1000$, at the level of 0.05 when $k = 1500$, and at the level of 0.1 When $k = 500$. It can be expected that the improvement of RCFS over SVM-RFE becomes less obvious when feature selection gains little from selecting too many or too few features. Even though, RCFS has never performed worse than SVM-RFE, in terms of number of wins and average test error.

Discriminative landmark selection. The following shows 3D landmark selection by RCFS, and the visual explanation of the obtained shape difference. Note that SVM-RFE cannot automatically deal with this problem. The landmark selection is conducted by selecting $k = 250$ and $k = 125$ landmarks respectively on 30 training and test groups for both left and right hippocampi. For $k = 250$, a linear SVM obtains the lowest test error rate 26.42% (left) and 24.53% (right) among the 30 test groups. For $k = 125$, the two lowest error rates

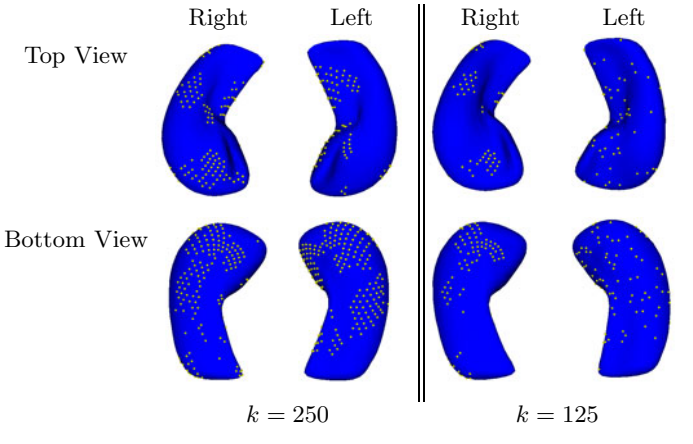


Fig. 1. Discriminative landmarks are selected in cases of $k = 250$ (left) and $k = 125$ (right) respectively. The selected landmarks are overlaid as the yellow balls on the mean shapes of the left and right hippocampi.

Table 4. Comparison of test error rates (%)of RCFS and NCFS for landmark selection

Test Error Rate (%)	left		right	
	$k = 250$	$k = 125$	$k = 250$	$k = 125$
NCFS	30.19	33.96	22.64	26.42
RCFS	26.42	26.42	24.53	22.64

becomes 26.42% and 22.64% respectively. The selected landmarks are overlaid on the mean shapes of the left and right hippocampi respectively, as shown in Fig. 1, to reveal the essential shape discrimination. By cross-referencing the results of $k = 250$ and $k = 125$, we can see that the majority of the identified differences locate in CA1 and subiculum surface zones, especially for the inferior part (bottom view). This observation agrees with some findings in the literature [6]. The sparsity of the selected landmarks is automatically determined by the RCFS algorithm. For example, the selected 125 landmarks of the left hippocampi are very sparse, while those in other cases are visually more gathered. However, as shown in Table 4, compared with NCFS where no redundancy constraints are imposed, RCFS achieves clearly better classification performance, except for the right hippocampi when 250 landmarks are selected. This demonstrates the advantage of RCFS.

4 Conclusion

SVM-RFE has been a fairly standard feature selection method used in many research fields. In this paper, we propose a constrained feature selection method that shows superior selection performance over SVM-RFE when noisy and redundant features exist. We apply it to identifying essential hippocampal shape difference between AD and the control. The proposed method can be efficiently solved as we carefully design the constraints and preserve its global solvability.

References

1. Vemuri, P., Gunter, J., et al.: Alzheimer's disease diagnosis in individual subjects using structural mr images: Validation studies. Neuroimage 39, 1186–1197 (2008)
2. Wang, L.: Feature selection with kernel class separability. IEEE Trans. Pattern Analysis and Machine Intelligence 30(9), 1534–1546 (2008)
3. Zhou, L., Wang, L., Shen, C.,: Feature selection with redundancy-constrained class separability. IEEE Trans. Neural. Networks 21(5), 853–858 (2010)
4. Dinkelbach, W.: On nonlinear fractional programming. Management Science 13(7) (1967)
5. Schrijver, A.: Theory of Linear and Integer Programming. John Wiley and Sons, Chichester (1986)
6. Csernansky, J., Wang, L., Swank, J., Miller, J., Gado, M., McKeel, D., Miller, M., Morris, J.: Preclinical detection of Alzheimer's disease: hippocampal shape and volume predict dementia onset in the elderly. Neuroimage 25, 783–792 (2005)

Simulation of Brain Mass Effect with an Arbitrary Lagrangian and Eulerian FEM

Yasheng Chen[1], Songbai Ji[2], Xunlei Wu[3], Hongyu An[1], Hongtu Zhu[3], Dinggang Shen[1], and Weili Lin[1]

[1] Dept. of Radiology, [3] Biostatistics Univ. of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA
[2] Thayer School of Engineering, Dartmouth Collage, Hanover, NH 03755, USA
[3] RENCI, 100 Europa Drive Suite 540 Chapel Hill, NC 27517, USA
{yasheng_chen,hongyuan,dgshen,weili_lin}@med.unc.edu, Songbai.Ji@Dartmouth.edu, xunlei@renci.org, hzhu@bios.unc.edu

Abstract. Estimation of intracranial stress distribution caused by mass effect is critical to the management of hemorrhagic stroke or brain tumor patients, who may suffer severe secondary brain injury from brain tissue compression. Coupling with physiological parameters that are readily available using MRI, eg, tissue perfusion, a non-invasive, quantitative and regional estimation of intracranial stress distribution could offer a better understanding of brain tissue's reaction under mass effect. A quantitative and sound measurement serving this particular purpose remains elusive due to multiple challenges associated with biomechanical modeling of the brain. One such challenge for the conventional Lagrangian frame based finite element method (LFEM) is that the mesh distortion resulted from the expansion of the mass effects can terminate the simulation prematurely before the desired pressure loading is achieved. In this work, we adopted an arbitrary Lagrangian and Eulerian FEM method (ALEF) with explicit dynamic solutions to simulate the expansion of brain mass effects caused by a pressure loading. This approach consists of three phases: 1) a Lagrangian phase to deform mesh like LFEM, 2) a mesh smoothing phase to reduce mesh distortion, and 3) an Eulerian phase to map the state variables from the old mesh to the smoothed one. In 2D simulations with simulated geometries, this approach is able to model substantially larger deformations compared to LFEM. We further applied this approach to a simulation with 3D real brain geometry to quantify the distribution of von Mises stress within the brain.

Keywords: ALEF, brain mechanics, mass effect, intracranial stress estimation.

1 Introduction

Brain mass effect caused by hemorrhagic stroke or uncurbed tumor growth may cause irreversible and life-threatening secondary brain injury. Because the intracranial space is confined by the skull, the accumulation of extra mass will compress brain

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 274–281, 2010.

tissue leading to elevated intracranial pressure, which will further cause downstream effects to reduce oxygen supply to cerebral tissue.

Image registration techniques have been exploited for estimating brain mass effect but results are limited. Most current image registration techniques, such as [1-3] do not fully consider the mechanical factors associated with deformation. As a result, the derived geometrical deformation may not truthfully reflect the underlying tissue deformation caused by the mass effect.

In contrast to these image warping methods, FEM biomechanical approaches in combination with medical image processing are promising for our goal. Some previously reported applications of FEM in medical imaging included capturing brain shift in open cranial surgery [4, 5] and modeling tumor growth inside the brain [6, 7]. From the FEM point of view, these approaches are categorized as either Lagrangian frame based FEM (LFEM) [4, 5] or Eulerian frame based FEM (EFEM) [6, 7], depending on how mesh is handled w.r.t. the modeled object. In LFEM, the mesh nodes are fixed on the object and following the same material points throughout deformation. In EFEM, the mesh is separated from the object and remains fixed in spatial domain without deforming with the object. Both LFEM and EFEM have weaknesses in modeling brain mass effect. A direct application of LFEM in simulating expansion of a mass region is difficult, because the large deformation from expansion of the mass region will distort the mesh and terminate the analysis prematurely. In order to solve this problem, an off-line remeshing procedure is necessary to generate a "fresh" mesh from the deformed geometry to continue the analysis. As a result, a complex mapping of the state variables (displacement and velocity fields) is required to maintain the continuity in analysis. In order to overcome the limitation of mesh distortion in LFEM, Hogea and colleagues have adopted EFEM to simulate large deformation in the brain. EFEM is able to handle extremely large deformation by separating mesh from the object. Because this approach fixes the mesh within the spatial domain, there may be multiple types of tissues within one grid. Thus, EFEM imposes an inherent difficulty in boundary handling (including boundary tracking, contacting problem, and boundary conditions). This well-known limitation makes EFEM an undesirable method to handle the potential contact problems caused by mass effect, which may include obstruction of aqueduct, sulci collapse or tissue-skull contact occurs. Furthermore, this work [6] was developed to facilitate image registration between tumor patient images with a normal template and no mechanical parametric results (e.g. strain or stress) have been given. In a later comparison work between EFEM and LFEM, Zacharaki *et al* demonstrated that similar results can be obtained with LFEM in Abaqus and the EFEM approach [7]. Thus, more justifications are needed for the utility of EFEM in modeling brain mass effect, particularly because EFEM was designed for modeling fluid dynamics, which has a different nature from solid mechanics.

In this work, we propose to simulate brain deformation caused by mass effect with an arbitrary Lagrangian Eulerian method based FEM (ALEF) [8]. This method was developed to combine the advantages of LFEM and EFEM. This algorithm consists

of three phases: 1) a Lagrangian phase to deform the mesh (similarly to LFEM), 2) a smoothing phase to reduce mesh distortion, and 3) an Eulerian phase to map the state variables to the new mesh. Compared to LFEM, ALEM reduces mesh distortion with its inherent mesh smoothing capability. Compared to EFEM, ALEF allows for boundary tracking by limiting the mesh smoothing within the tissue boundary. In this work, we will evaluate the application of this approach in simulating brain tissue deformation caused by the expansion of a mass region in both simulated and real brain geometries. We will demonstrate that compared to LFEM, ALEF can simulate substantially larger deformation caused by expansion of the mass region.

2 Methods

2.1 Geometrically Nonlinear LFEM

In LFEM, strain tensor matrix is computed via Eq. (1). In this approach, the material points from the original (un-deformed) configuration (with coordinates ${}^{0}x_i$ at time *0* in Eq. (1)) are tracked throughout the analysis (with coordinates ${}^{t}x_i$ at time *t*).

$$ {}_{0}^{t}X_{i,j} = \partial {}^{t}x_i / \partial {}^{0}x_i \,,\; {}_{0}^{t}C = {}_{0}^{t}X^{T} * {}_{0}^{t}X \,,\; {}_{0}^{t}E = \tfrac{1}{2}({}_{0}^{t}C - I) \tag{1} $$

Usually, when the strain is less than 10%, the higher order terms (i.e. 2nd order) in *C* and *E* (strain tensor matrix) are negligible, and geometrically linear FEM can be used for analysis. But in our application, the large deformation from expansion of the mass region will result in strain values well above 10%, and these high order terms are preserved for a more accurate nonlinear simulation. From the virtual work principle, the equilibrium equation is given in Eq. (2)

$$ \int_{{}^{0}V} {}_{0}^{t}S_{i,j}\,\delta\, {}_{0}^{t}E_{i,j}\, d\,{}^{0}V + \int_{{}^{0}V} \rho_i \ddot{u}_i\, \delta u_i\, d\,{}^{0}V = \int_{{}^{0}V} {}_{0}^{t}f_i^{B}\, \delta u_i\, d\,{}^{0}V + \int_{{}^{0}S} {}_{0}^{t}f_i^{S}\, \delta u_i^{S}\, d\,{}^{0}S \tag{2} $$

where, the total work done by the external force (including body force ${}_{0}^{t}f_i^{B}$ and surface force ${}_{0}^{t}f_i^{S}$) equals the total sum of the energy deposited within the continuum through deformation (1st term in the left-hand-side of Eq. (2)) and acceleration (2nd term in the left-hand-side of Eq. (2)). After matrix assembly, the matrix form used for solution was derived as in the format of Eq. (3). In our work, we chose explicit dynamics to solve the equation due to the dynamic feature of brain mass effect. Different speed of mass accumulation or pressure development inside the mass region will have different pathological implications. One such example is that brain tissue's reaction towards the accumulation of mass may be different between hemorrhagic stroke and brain tumor patients. The static solution neglecting the acceleration and speed related force terms in Eq. (3) does not fit for our purpose.

$$ M\ddot{U} + D\dot{U} + K(U)U = R \tag{3} $$

2.2 Arbitrary Lagrangian Eulerian Method

ALE method involves a total of three frames, material frame(X), spatial frame (x), and the mesh frame (m, a.k.a. reference frame). The mappings χ and ψ represent the mappings from the mesh frame towards the material and spatial frames respectively. Applying chain rules in connecting these mappings, the physical velocity ($v(m,t)$), referential velocity ($v_m(m,t)$), convective velocity ($c(m,t)$) are derived as in Eq. (4).

$$v(m,t) \coloneqq d\psi(\chi^{-1}(X,t),t)/dt = \left.\frac{\partial x}{\partial t}\right|_m + grad_m x \left.\frac{\partial m}{\partial t}\right|_X \qquad v_m(m,t) \coloneqq \frac{\partial \psi(m,t)}{\partial t} = \left.\frac{\partial x}{\partial t}\right|_m$$
$$c(m,t) = v(m,t) - v_m(m,t) = grad_m x \left.\frac{\partial m}{\partial t}\right|_X \tag{4}$$

Similarly, the conservative equations for mass, linear momentum and energy are derived as in Eq. (5), where ρ, v, b, u and σ respectively stand for tissue density, velocity in spatial domain, body force, energy term and stress.

$$\left.\frac{d\rho}{dt}\right|_m + grad_x \rho \bullet c + \nabla \cdot v = 0 \quad \rho\left(\left.\frac{\partial v}{\partial t}\right|_m + grad_x v \bullet c\right) = div_x \cdot \sigma + \rho b$$
$$\rho\left(\left.\frac{\partial u}{\partial t}\right|_m + c \cdot \nabla e\right) = \sigma : \nabla^S v \tag{5}$$

2.3 Mesh Smoothing

The central idea of ALEF is to include an inline smoothing phase to partially seperate the mesh from the object and allow the mesh to move within the regions of the same tissue property to reduce distortion. In this approach, due to its robustness, volume based smoothing was employed as in Eq. (6), where Ve represents the volume of element, e, which is neiboring to node i (node i is one node of element e) with the centroid location x_e. For a node inside one material region, the volume based smoothing automatically guarantees the smoothed location remains inside the same material region. For a boundary node, if it is co-planar with its neighboring nodes, it is allowed to move on the plane; otherwise, it remains fixed.

$$x_i = \sum_{e=1}^{N} V_e x_e \Big/ \sum_{e=1}^{N} V_e \tag{6}$$

2.4 Uncoupled Solution

It is apparent that in Eq. (5) all the conservative equations assume a similar structure. Thus, operator splitting was used to break the linear advection equations in Eq. (5) into two simpler forms as in Eq. (7), where ϕ represents ρ, v, u terms in Eq. (5).

$$\frac{\partial \phi}{\partial t} + c\frac{\partial \phi}{\partial x} = f \quad \phi(x,0) = \phi_0(x) \;\text{->}\; \frac{\partial \phi}{\partial t} = f \quad \frac{\partial \phi}{\partial t} + c\frac{\partial \phi}{\partial x} = 0 \tag{7}$$

We obtain the solution of the Lagrangian phase through Taylor expansion (Eq. (8)).

$$\phi_{n+1}^{L} = \phi_n + \left.\frac{\partial \phi_n}{\partial t}\right|_X \cdot \Delta t \quad \text{with} \quad \left.\frac{\partial \phi_n}{\partial t}\right|_X = f \tag{8}$$

After Lagrangian phase, the mesh smoothing (Eq. (6)) were applied to the current deformed mesh to reduce distortion (multiple sweeps can be applied if necessary). The convective velocity field c was computed through the relative motion of the nodes of mesh before and after smoothing. The mapping from the old mesh to the smoothed mesh is computed as a second order advection through a flux-limiting method [9]. Due to the equivalence between the spatial and temporal derivatives (the splitted PDE in Eq. (7)), the time based updating in this Eulerian phase can be computed through spatial derivatives (Eq. (9)).

$$\phi_{n+1} = \phi_{n+1}^{L} + \left.\frac{\partial \phi_{n+1}^{L}}{\partial t}\right|_m \cdot \Delta t + \frac{1}{2}\left.\frac{\partial^2 \phi_{n+1}^{L}}{\partial t^2}\right|_m \cdot \Delta t^2 \qquad \left.\frac{\partial \phi_{n+1}^{L}}{\partial t}\right|_m = -c_i \frac{\partial \phi_{n+1}^{L}}{\partial x_i} \quad \left.\frac{\partial^2 \phi_{n+1}^{L}}{\partial t^2}\right|_m = c_i c_j \frac{\partial^2 \phi_{n+1}^{L}}{\partial x_i \partial x_j} \tag{9}$$

3 Results

3.1 Simulation with Homogeneous Geometry

We first evaluated the performances of ALEF and LFEM in a simulation using a simplified geometry consisting of one homogeneous material with a Young's modulus (YM) = 2000pa and Poisson's ratio (PR) = 0.45, similar to the white matter parameters used in [6]. A small off-center seed circle was inflated with three pressure loadings smoothly increased from 0pa to 800pa, 1000pa and 1200pa in 1000 seconds.

The simulation was performed with Abaqus (SIMULIA, Rhode Island). When the pressure loading was low (800pa), both LFEM and ALEF produced similar results (2nd column, Fig. 1). When pressure was increased to 1000pa and 1200pa, only ALEF was able to complete the inflation process successfully. LFEM terminated prematurely and resulted in less expansion of the mass region compared to ALEF. The areas of the final mass region obtained with LFEM were respectively 55% (1000pa) and 38% (1200pa) of the final areas obtained with ALEF (apparent in the areas of the white mass regions in Fig. 1.)

3.2 Simulation with Multi-shelled Geometry

In this simulation, a more complex geometry consisting of five shells representing skull, gray matter surface, white matter surface, ventricle and the mass region. Parameters from linear elastic models as in [6] are used (white matter, YM=2000pa, PR=0.45; gray matter, YM=2500pa, PR=0.45; CSF, YM=500pa, PR=0.1). Three simulations were performed with the center of the mass region initialized as the center of white matter region (Case 1), 5mm closer to the ventricle (Case 2), and 5mm closer to the grey matter (Case 3) from the center location. In all the three cases, the mass regions were inflated with a pressure loading smoothly increased from 0pa to 1000pa in 1000 seconds. As in previous simulation, in all the three cases, ALEF is able to

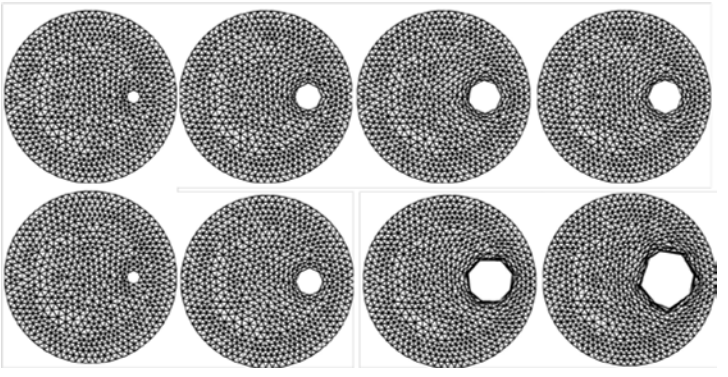


Fig. 1. The initial un-deformed geometry (1st column) and the simulation results with pressure loadings: 800pa (2nd column), 1000pa (3rd column) and 1200pa (4th column) in both LFEM (top row) and ALEF (bottom row)

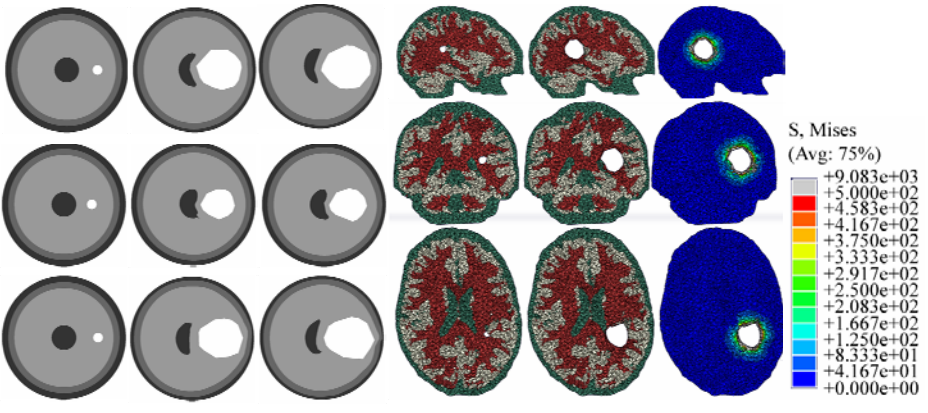


Fig. 2. (a) Left panel. Simulation results with the initial mass region located at the center of the white matter region (1st row), 5mm close to the ventricle (2nd row) and 5mm closer towards gray matter (last row) from LFEM (2nd column) and ALEF (3rd column). (b) Right panel. The simulation with a real brain geometry with initial seeded sphere (1st column), after deformation(2nd column) and distribution of von Mises stress(last column).

produce a larger deformation compared to LFEM. The final area of the mass region obtained LFEM are respectively 85%, 85% and 87% of the results of ALEF (apparent as the larger white areas in ALEF in Fig. 2a).

3.3 Estimating von Mises Stress Distribution with Real Brain Geometry

In this simulation, an artificial sphere as a seed region for mass accumulation was planted inside a real 3D brain geometry obtained from MRI (1st column, Fig. 2b). The mechanical parameters as in the previous simulation were used. The pressure loading was 800Pa. The von Mises stress was computed from the stress tensor matrix for

each element (Eq. (10)). For linear elastic material, the material starts to yield when the von Mises stress reaches a critical value (i.e. yield point). In such a situation, the tissue deformation may become irreversible. Our study demonstrated that at early stage of mass accumulation, von Mises stress descended steeply from the immediate peri-mass region to distal area (Fig. 2b).

$$\sigma_v = \sqrt{((\sigma_{11}-\sigma_{22})^2+(\sigma_{11}-\sigma_{33})^2+(\sigma_{22}-\sigma_{33})^2+6(\sigma_{12}^2+\sigma_{13}^2+\sigma_{23}^2))/2} \quad (10)$$

4 Discussion

In this study, we adopted ALEF with explicit dynamics to simulate brain deformation caused by mass effect. We have demonstrated that this approach is more appropriate for analysis of large deformation compared to LFEM. This is critical, since LFEM may significantly underestimate the expansion of the mass region. In this work, the comparisons were performed with the simulation results obtained with Abaqus. A more ideal testing with analytical solutions in simulated 2D geometries as the ground truth may be more convincing. But given the popularity of Abaqus in both industrial and academic environments, we expect that the conclusion holds even when analytical solutions are used as ground truth.

The improvement in performance of ALEF is not necessarily achieved at the cost of prolonged computational time. On the contrary, ALEF may reduce analysis run-time significantly compared to LFEM. This is because the critical time interval for updating the partial differential equations is proportional to the smallest characteristic length of the element in the mesh. In LFEM, the reduction in characteristic length caused by mesh distortion will dramatically reduce step size for updating the equations. Sometimes convergence becomes impossible and simulation may terminate prematurely. In contrast, ALEF can reduce mesh distortion and help the system to maintain a larger time step size for faster dynamic simulation.

Our study may be one of the earliest studies which are able to provide stress distribution within brain due to mass effect based upon subject-specific brain geometry. Since our approach is based on explicit dynamics, it is possible to predict the progression of brain mass effect and allow for early intervention. Furthermore, combining with advanced MRI techniques such as perfusion and oxygenation measurement, a more complete picture of how brain tissue reacts under mass effect can be revealed.

There are a few limitations to our current study. ALEF's capability to reduce mesh distortion is not effective without a limit. In extremely large deformation (e.g. hundreds of times increase in volume of the mass region), this inline smoothing in ALEF may not be able to tackle mesh distortion completely. In this case, an off-line re-meshing program can be executed to obtain a fresh mesh to resume the simulation. The pressure loading used for testing was uniformly distributed across mass region, and this situation may not hold in certain diseased conditions. Finally, our analyses were performed under the assumption that brain tissue behaves like linear elastic materials. However, ALEF can be adapted to model nonlinear material models such as hyper-elasticity [10] with ease. In this case, the Lagrangian phase of ALEF needs to compute the stress as a function of strain rate with the hyper-elastic material.

Acknowledgments. This work is supported in part by AHA grant 0730321 to Dr. An, NSF grant BCS-08-26844, NIH grants UL1-RR025747-01, MH086633, P01CA142538-01 and AG033387 to Dr. Zhu.

References

1. Shen, D., Davatziko, D.: HAMMER: Hierarchical Attribute Matching Mechanism for Elastic Registration. IEEE-TMI 21(11), 1421–1439 (2002)
2. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. Int. J. Comput. Vis. 61(2), 139–157 (2005)
3. Clatz, O., Delingette, H., Talos, I., Golby, A., Kikinis, R., Jolesz, F., Ayache, N., Warfield, S.: Robust Nonrigid Registration to Capture Brain Shift From Intraoperative MRI. IEEE-TMI 24(11), 1334–1346 (2005)
4. Miga, M.I., Paulsen, K.D., Lemery, J.M., Eisner, S., Hartov, A., Kennedy, F.E., Roberts, D.W.: Model-Updated Image Guidance: Initial Clinical Experiences with Gravity-Induced Brain Deformation. IEEE-TMI 18, 866–874 (1999)
5. Wittek, A., Miller, K., Kikinis, R., Warfield, S.K.: Patient-Specific Model of Brain Deformation: Application to Medical Image Registration. Journal of Biomechanics 40, 919–929 (2007)
6. Hogea, C., Biros, G., Abraham, F., Davatzikos, C.: A Robust Framework for Soft Tissue Simulations with Application to Modeling Brain Tumor Mass Effect in 3D MR images. Physics in Medicine and Biology 52, 6893–6908 (2007)
7. Zacharaki, E.I., Hogea, C.S., Biros, G., Davatzikos, C.: A Comparative Study of Biomechanical Simulators in Deformable Registration of Brain Tumor Images. IEEE-BME 55(3), 1233–1236 (2008)
8. Belytschko, T., Liu, W.K., Moran, B.: Nonlinear Finite Elements for Continua and Structures. John Wiley & Sons, Chichester (2002)
9. Van Leer, B.: Towards the Ultimate Conservative Difference Scheme. A New Approach to Numerical Convection. Journal of Computational Physics 23, 276–299 (1977)
10. Miller, K., Chinzei, K.: Mechanical Properties of Brain Tissue in Tension. J. Biomechanics 35(4), 483–490 (2002)

Relating Structural and Functional Connectivity to Performance in a Communication Task

Jeffrey T. Duda[1], Corey McMillan[2], Murray Grossman[2], and James C. Gee[3]

[1] Department of Bioengineering, University of Pennsylvania, USA
jtduda@seas.upenn.edu
[2] Department of Neurology, University of Pennsylvania, USA
[3] Department of Radiology, University of Pennsylvania, USA

Abstract. Measures from event-related functional MRI, diffusion tensor imaging tractography and cognitive performance in a language-based task were used to test the hypothesis that both functional and structural connectivity provide independent and complementary information that aids in the identification of network components most related to the neurobiological basis for language and cognitive processing. Structural connectivity was measured by averaging fractional anisotropy (FA) over a geometric fiber bundle model that projects local white matter properties onto a centerline. In the uncinate fasciculus FA was found to predict performance on a measure of decision-making regarding homonym meaning. Functional synchronization of BOLD fMRI signals between frontal and temporal regions connected by the uncinate fasciculus was also found to predict the performance measure. Multiple regression analysis demonstrated that combining equidimensional measures of functional and structural connectivity identified the network components that most significantly predict performance.

1 Introduction

In this paper functional subnetworks in the brain are examined using MRI to measure both structural connectivity (SC) and functional connectivity (FC). Additionally, the influence on behavior of both SC and FC is examined to determine the degree to which each provides unique information as well as how this information may be used to identify the parts of a network that are most influential on behavioral performance. FC involves co-activation of brain regions during performance of a task while brain recruitment is monitored with fMRI. SC is related to the long tract white matter projections that may integrate recruited brain regions biologically. The use of diffusion tensor imaging (DTI) in tract-specific studies has received much attention recently and has demonstrated the utility of atlas-based methodologies [1,2]. Geometric models of fiber bundles are typically used to create arc-length parameterizations of diffusion properties [3,4,5]. Statistical analysis of diffusion properties of fiber bundles may be enhanced by reducing dimensionality by projecting local properties onto a skeleton [2] or medial-axis [1]. Inspired by the success of these approaches, we

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 282–289, 2010.

developed an atlas-based approach for using diffusion tensor tractography to determine geometric models of fiber bundles that connect functionally defined cortical regions of interest. Combining FC and SC, and further integrating behavioral data, provides important insight into the nature of the relationship between structure and function in the brain and their respective roles in determining behavior [6,7,8]. In this study, we demonstrate how SC and FC may be used to examine small, functionally defined subnetworks in the brain during performance of a common language task. Functionally defined cortical regions are used along with a population-averaged diffusion tensor atlas to identify the white matter pathways that provide the basis for biological connectivity. A centerline-based method is used to provide a geometric model that facilitates the equidimensional comparison of FC and SC within a network. Behavioral data are used to identify the relative contributions of function and structure, and the degree to which each provides unique insight into behavior.

2 Methods

Functionally defined cortical regions and the white matter pathways that connect them are used to measure both structural and functional connectivity in individual subjects. Multiple regression analysis of cognitive performance scores is then used to determine the components of the subnetworks that most influence performance and the relative predictive strengths of the connectivity measures. A set of ten subjects (6 female, mean age of 23.5 years) had T1, functional and diffusion tensor images acquired. These are used to examine the relationship between structure and function and their relationship to performance in the functional task.

2.1 Subnetworks in the Brain

We investigate a hypothesis that suggests language processing is supported by the interaction of two cortical subnetworks [9]. This hypothesis specifically focused on the strategic process of minimizing ambiguity during language production, defined as an individuals use of an unambiguous work (e.g. "cage") instead of a semantically ambiguous word (e.g. "pen") to make the meaning of an utterance more clear. The cortical regions and MNI coordinates of peak activation are listed in figure 1. A forced choice paradigm in which subjects were instructed to make the choice that resulted in the most clear sentence meaning (e.g. reduced ambiguity) was used. The proportion of unambiguous choices was used as the measure of cognitive performance. It has been proposed that this process of minimizing ambiguity relies on the recruitment of two subnetworks: a lexical semantic network situated in posterolateral temporal cortex (PLC) and anterior temporal cortex (ATC) to support retrieval of a word form; a strategic decision-making network to support probabilistic resources in the dorsolateral prefrontal cortex (DLPFC) (e.g. evaluating the likelihood that a word like "pen" has multiple meanings), risk-based resources in orbital frontal cortex (OFC) (e.g. using

"pen" in an ambiguous context has a higher "risk" of being misinterpreted), and an integration mechanism in inferior parietal cortex (IPC) that brings together probabilistic and risk information to inform a decision. To evaluate the biological feasibility of these networks, the current work evaluated the white matter connections between the activated regions and measured FC between all regions.

2.2 Analysis of fMRI

For the examination of FC the previously cortical regions were examined. A region-growing process was used to create 5 regions of equal volume. Each peak activation coordinate was used as a seed for the region growing algorithm in which each region was grown iteratively by examining neighboring voxels in the region and adding each of the 8-connected neighbors in a template-based brain mask created in MNI space. The process iterated until each region had a volume of $480mm^3$. In each subject, an average time-course BOLD fMRI signal was calculated for each activated region. To estimate FC between regions, a Pearson's correlation coefficient was calculated using each regions' averaged signal.

2.3 Analysis of DTI

Diffusion tensor tractography in individual subjects is highly subject to false-positive connections, but recent work has shown that the improved SNR provided by a population atlas provides an appropriate space for identifying fiber bundle geometry [10]. To achieve this, a multivariate atlas was created from a data set consisting of 26 healthy young adults for whom both high resolution T1 images and DTI were acquired. The 10 subjects examined in this study were all included in this atlas-building data set. The set of all subjects' high resolution T1 weighted images were registered to the template using Symmetric Normalization as implemented in Advanced Normalization Tools [11]. This was accomplished through the use of a multi-resolution, non-rigid registration algorithm to optimize a cross correlation metric under the constraints of a diffeomorphic transformation model [11]. A brain mask of the template was propagated to each subject's T1 weighted image. These skull-stripped T1 weighted images were then registered to the FA image derived from each subject's diffusion tensor image. The intra-subject transforms were composed with the T1 atlas transforms in order to transform the each subjects' tensor data into template space using the preservation of principle technique along with log-Euclidean linear interpolation.

The diffusion tensor component of the atlas was used to perform whole brain, deterministic fiber tractography [12]. Landmarks were manually placed in the T1 component of the atlas in order to extract well defined white matter fiber bundles [13]. The functionally activated regions were warped from MNI space into the template space and dilated by 5mm to extend into the white matter for use as target regions to identify fiber bundles that connected two regions of interest. A popular approach to avoiding partial voluming bias is to incorporate a centerline or skeletonization technique in which an FA value at each point is determined by finding the maximum FA in the local neighborhood [2,1]. Here

we use a template-fiber approach where an elliptical cross-section is defined at each point along the centerline and used to identify local maxima. For each white matter tract, the atlas-based tractography provided a bundle of streamlines, each of which was parametrized by arc-length to extend from 0.0 to 1.0. A BSpline was then fit to the set of all points from all streamlines in each bundle to obtain a single centerline that lies in the core of the fiber pathway of interest. For each point along the model pathway, a tangent was calculated and used to determine a perpendicular plane. The intersection of this plane with each of the streamlines in the bundle defines a set of points. The normal and binormal vectors were used to re-parameterize the intersection points into 2D coordinates. Graham's scan method was used to determine the convex hull that encloses the set of intersection points [14], and least-squared method was applied to the points on the hull to define an elliptical cross-section [15]. These model pathways were then used to examine each subject's FA image. At each point along the average pathway, the maximum fractional anisotropy within the associated elliptical cross-section was projected to the centerline. To obtain a single SC value for the entire fiber bundle, the FA values were averaged over the length of the centerline.

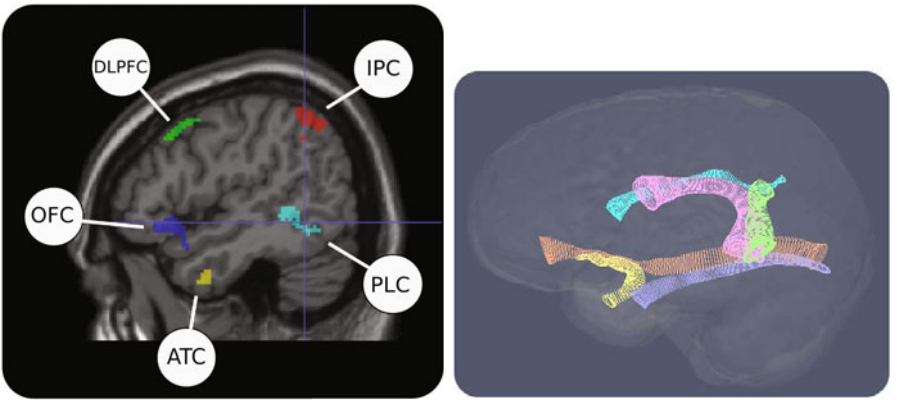
2.4 Relating Structure and Function to Behavior

Behavioral data was used to explore the extent to which FC and SC provide unique and relevant information regarding behavior. The proportion of responses in which subjects minimized ambiguity was used to identify the components of the network that most directly influence performance in the functional task. Multiple regression analysis was performed to examine the relationship between FC and SC between regions for which there exists a direct white matter connection. Both FC and SC were used as independent variables to quantify their relationship to the behavioral scores.

3 Results

The functionally activated regions illustrated in figure 1 were used to identify the white matter tracts of interest revealing a biological network made up of the uncinate fasciculus (UF), arcuate fasciculus (AF), superior longitudinal fasciculus (SLF), inferior longitudinal fasciculus (ILF), inferior frontal-occipital fasciculus (IFO) and an inferior-superior fiber bundle running along the arcuate fasciculus that will be referred to as the vertical aspect of the arcuate fasciculus (AFv). The elliptical cross-sections used to define local maxima along each template fiber are illustrated in figure 1 and are consistent with known neuroanatomy. The results of the SC and FC measurements are summarized in table 1. The posterior lateral temporal cortex (PLC) had the highest average FC values. This region also has the highest number of white matter connections as it directly connects to every other cortical region in the network.

To facilitate the comparison of structure and function, the functional analysis was limited to examining FC between regions that were determined to have direct



(A) (B)

Fig. 1. A) Functionally activated cortical regions were used to determine an MNI coordinate of peak activation in posterior lateral temporal cortex (PLC=-64 x -38 x 6), anterior temporal cortex (ATC=-54 x 22 x -6), dorso-lateral prefrontal cortex (DLPFC=-50 x 20 x 42), inferior parietal cortex (IPC=-32 x -64 x 58) and orbital frontal (OFC=-54 x 2 x -34) B) The cortical regions were used to identify white matter fiber bundles that connected any two regions of interest. This resulted in the identification of the uncinate fasciculus (yellow), arcuate fasciculus (pink), orbito frontal-occipital fasciculus (orange), inferior longitudinal fasciculus (purple), superior longitudinal fasciculus (light blue) and a vertical aspect of the arcuate fasciculus (green)

biological connections. For each participant, a behavioral score was determined by the ratio of times that they correctly chose the unambiguous alternative. The mean score was 0.717 ± 0.14. For each white matter tract the FC and SC values were used as independent variables in multiple linear regression on the behavior scores, using R. The resulting p-values, summarized in figure 2, were FDR corrected and connectivity in the UF was found to be most significantly correlated ($p = 0.014, r^2 = 0.810$) to performance. In order to examine relative strengths, SC and FC are standardized against one another to calculate their beta coefficients which gives $\beta_{FC} = 0.425$ and $\beta_{SC} = 0.681$. To illustrate the

Table 1. Functional connectivity was measured using a Pearson's correlation for each pair of regions (upper triangular values) and structural connectivity was measured for each pair of regions that had a direct white matter connection by examining FA (lower triangular values)

	IPC	DLPFC	OFC	ATC	PLC
IPC	-	0.59 ± 0.29	0.61 ± 0.14	0.49 ± 0.26	0.49 ± 0.22
DLPFC	0.62 ± 0.07	-	0.27 ± 0.33	0.49 ± 0.19	0.69 ± 0.12
OFC	-	-	-	0.38 ± 0.36	0.68 ± 0.15
ATC	-	-	0.50 ± 0.10	-	0.60 ± 0.24
PLC	0.64 ± 0.08	0.66 ± 0.07	0.68 ± 0.05	0.60 ± 0.12	-

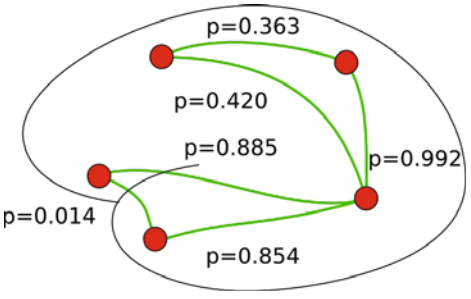
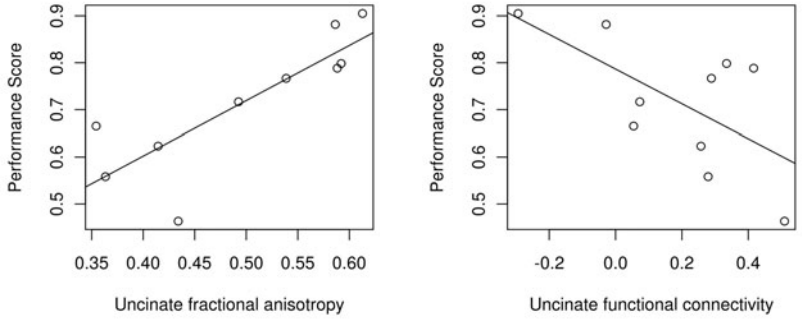


Fig. 2. Both structural and functional connectivity in each connection were used in a multiple linear regression analysis of performance. FDR-corrected p-values are shown here. Connectivity in the uncinate ($p = 0.014$) was most significantly correlated to performance.



(A) (B)

Fig. 3. Regression analysis of cognitive performance was examined independently to examine correlations with (A) structural connectivity and (B) functional connectivity for the uncinate fasciculus

unique contributions of each connectivity measure, both both FC and SC in the UF were independently used as independent variable in a regression analysis of performance. Correlating SC to performance resulted in $p = 0.003, r^2 = 0.657$ while FC to performance resulted in $p = 0.034, r^2 = 0.379$. These results are illustrated in figure 3. A regression analysis using SC to predict FC resulted in $p = 0.308, r^2 = 0.020$.

4 Discussion

This study demonstrated that SC and FC provide unique and converging evidence for identifying the components of a brain network that are most significantly related to cognitive performance. The use of geometric models for white

matter fiber bundles provided a robust framework for quantifying tract-wide metrics of structural integrity that directly correlated to a behavioral measure. FC between regions was quantified using the Pearson's correlation of average BOLD fMRI signals. Using these connectivity measures in a multiple regression analysis implicated the UF as being the network component that most directly influences performance. When examined independently, both SC and FC in the UF significantly correlated to performance, but the relatively weak correlation between SC and FC suggests that each provides unique information about network connections.

The identification of the UF as the network connection that most mediates performance is interesting as it provides the connection between a region commonly associated with language processing (ATC) and a region commonly implicated for decision-making (OFC). A connection between these regions supports the hypothesis that ambiguity minimization is supported by the interaction of two cortical subnetworks. The UF provides the only direct connection between these regions, but an indirect connection is provided by the ILF and IFO which both connect to the PLC, a region associated with language processing. While the UF is not the only direct connection between the OFC and a language processing region, it is the shortest connection, and is thus consistent with studies of both SC [16] and resting-state FC [17] that have revealed evidence suggesting that shorter connective distances are associated with higher connectivity.

The use of template fibers with elliptical cross-sections provided an effective geometric model for examining white matter fiber bundle properties, but a great deal of opportunity exists for the development of more biologically relevant measures of SC. Here, these models were purely template-based and used to examine FA in individual subjects. Using these template-based models as a basis for fitting subject-specific models from subject-space tractography could potentially provide a more sensitive measure of SC and could provide a framework that explicitly examines the geometry of white matter pathways as well as the properties of the underlying tissue. Additionally, the use of metrics that leverage the expected fiber orientation provided by the geometric model may be useful as they incorporate more widespread information about the fiber tract as opposed to the purely local measure provide by FA [18].

References

1. Yushkevich, P.: Structure-specific statistical mapping of white matter tracts. NeuroImage (2008)
2. 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.J.: Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31(4), 1487–1505 (2006)
3. Jones, D.K., Travis, A.R., Eden, G., Pierpaoli, C., Basser, P.J.: Pasta: pointwise assessment of streamline tractography attributes. Magn. Reson. Med. 53(6), 1462–1467 (2005)

4. Maddah, M., Grimson, W.E.L., Warfield, S.K., Wells, W.M.: A unified framework for clustering and quantitative analysis of white matter fiber tracts. Med. Image Anal. 12(2), 191–202 (2008)
5. O'Donnell, L.J., Westin, C.F., Golby, A.J.: Tract-based morphometry for white matter group analysis. Neuroimage 45(3), 832–844 (2009)
6. Baird, A.A., Colvin, M.K., Van Horn, J.D., Inati, S., Gazzaniga, M.S.: Functional connectivity: Integrating behavioral, diffuison tensor imaging, functional magnetic resonanace imaging data sets. Journal of Cognitive Neuroscience 17, 687–693 (2005)
7. Madden, D.J., Spaniol, J., Whiting, W.L., Bucur, B., Provenzale, J.M., Cabeza, R., White, L.E., Huettel, S.A.: Adult age differences in the functional neuroanatomy of visual attention: a combined fmri and dti study. Neurobiol. Aging 28(3), 459–476 (2007)
8. Stevens, M.C., Skudlarski, P., Pearlson, G.D., Calhoun, V.D.: Age-related cognitive gains are mediated by the effects of white matter development on brain network integration. Neuroimage 48(4), 738–746 (2009)
9. McMillan, C.T., Gunawardena, D., Ryant, N., Clark, R., Grossman, M.: Converging evidence for decision-making in successful communication. In: Proceedsings of Society for Neuroscience's 38th Annual Meeting, Chicago, USA (2009)
10. Goodlett, C.B., Fletcher, P.T., Gilmore, J.H., Gerig, G.: Group analysis of dti fiber tract statistics with application to neurodevelopment. Neuroimage 45(suppl. 1), S133–S142 (2009)
11. Avants, B.A., Schoenemann, P.T., Gee, J.C.: Lagrangian frame diffeomorphic image registration: Morphometric comparison of human and chimpanzee cortex. Medical Image Analysis 10(3), 397–412 (2006)
12. 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: Proceedings of ISMRM, vol. 14, p. 2759 (2006)
13. Wakana, S., Jiang, H., Nagae-Poetscher, L.M., van Zijl, P.C.M., Mori, S.: Fiber tract-based atlas of human white matter anatomy. Radiology 230(1), 77–87 (2004)
14. Graham, R.L.: An efficient algorithm for determining the convex hull of a finite planar set. Information Processing Letters 1, 344–370 (1972)
15. Fitzgibbon, A.W., Pilu, M., Fisher, R.B.: Direct least-squares fitting of ellipses. PAMI 21(5), 476–480 (1999)
16. Lewis, J.D., Theilmann, R.J., Sereno, M.I., Townsend, J.: The relation between connection length and degree of connectivity in young adults: a dti analysis. Cereb. Cortex 19(3), 554–562 (2009)
17. Honey, C.J., Sporns, O., Cammoun, L., Gigandet, X., Thiran, J.P., Meuli, R., Hagmann, P.: Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. USA 106(6), 2035–2040 (2009)
18. Duda, J., Avants, B., Asmuth, J., Zhang, H., Grossman, M., Gee, J.: A fiber tractography based examination of neurodegeneration on language-network neuroanatomy. In: Proc. Workshop on Computational Diffusion MRI (MICCAI 2008), New York, NY (2008)

Bayesian Classification of Multiple Sclerosis Lesions in Longitudinal MRI Using Subtraction Images*

Colm Elliott[1], Simon J. Francis[2], Douglas L. Arnold[3], D. Louis Collins[2], and Tal Arbel[1]

[1] Centre for Intelligent Machines, McGill University, Canada
[2] Montreal Neurological Institute, McGill University, Canada
[3] NeuroRx Research, Montreal, Canada

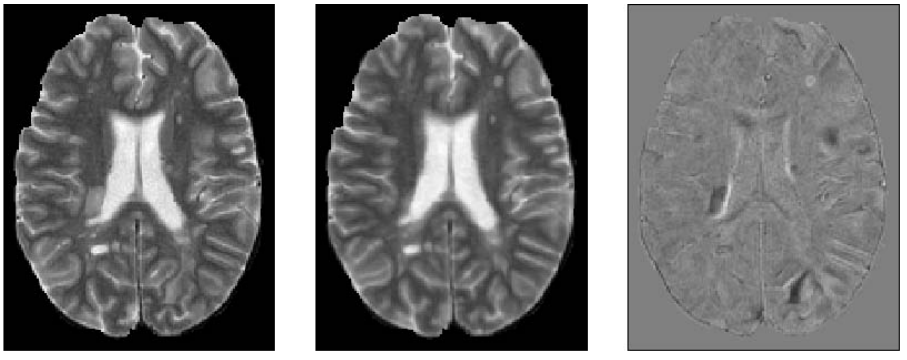
Abstract. Accurate and precise identification of multiple sclerosis (MS) lesions in longitudinal MRI is important for monitoring disease progression and for assessing treatment effects. We present a probabilistic framework to automatically detect new, enlarging and resolving lesions in longitudinal scans of MS patients based on multimodal subtraction magnetic resonance (MR) images. Our Bayesian framework overcomes registration artifact by explicitly modeling the variability in the difference images, the tissue transitions, and the neighbourhood classes in the form of likelihoods, and by embedding a classification of a reference scan as a prior. Our method was evaluated on (a) a scan-rescan data set consisting of 3 MS patients and (b) a multicenter clinical data set consisting of 212 scans from 89 RRMS (relapsing-remitting MS) patients. The proposed method is shown to identify MS lesions in longitudinal MRI with a high degree of precision while remaining sensitive to lesion activity.

1 Introduction

The use of subtraction imaging to identify MS lesion activity on MRI has been shown to increase sensitivity to new and resolving lesions and significantly reduce inter-rater variability [1,2,3,4]. Previous studies using subtraction images for lesion identification were done in a manual or semi-automatic fashion, as the automatic analysis of subtraction images is complicated by the presence of registration errors, flow artifacts and the high variability of signal intensities for lesions [3,5]. For this reason, most automated longitudinal MS lesion segmentation approaches have used more robust statistical approaches that require the inclusion of images from several timepoints [6,7,8], the analysis of lower resolution patches [9], or some form of deformation analysis [5,10]. The automatic lesion classifier presented here attempts to overcome the limitations of subtraction imaging by embedding intensity differences into a Bayesian framework that also incorporates prior classification at a reference timepoint, models that account for registration error and noise, and neighbourhood information. Our probabilistic

* This work was supported by NSERC Strategic Grant (350547-07).

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 290–297, 2010.



(a) (b) (c)

Fig. 1. T2 images for 2 timepoints are shown in (a) and (b) while the subtraction image between the two timepoints is shown in (c). New lesions appear as hyperintense on the T2 subtraction image while resolving (disappearing) lesions appear as hypointense.

framework further permits qualification of the degree of confidence to classification results at each voxel in the form of a posterior probability for each tissue class.

Our method was evaluated on (a) a scan-rescan data set consisting of 3 MS patients and (b) a multicenter clinical data set consisting of 212 scans from 89 RRMS patients with 2-4 longitudinal scans each. The overall classification system provides a consistent labelling of lesion voxels while remaining sensitive to lesion activity.

2 Methods

2.1 Problem Formulation

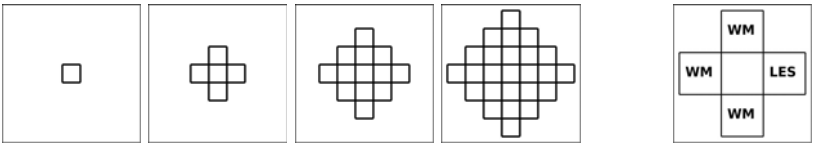
We present the problem of classification as one of inferring a tissue class label, $C_i^{(t)}$, at each voxel i of a multimodal volume at timepoint t, given an image from a reference timepoint, $I^{(r)}$, and a subtraction image, $D^{(t)}$, between timepoint t and the reference timepoint. Tissue class labels are restricted to one of cerebrospinal fluid (csf), gray matter (gm), white matter (wm), MS lesion (les) and a partial volume class (pv).

We first formulate the tissue class inference problem by only considering observations at the voxel in question:

$$p_i^0 = p(C_i^{(t)}|I_i^{(r)}, D_i^{(t)}) = \sum_{C_i^{(r)}} p(C_i^{(r)}, C_i^{(t)}|I_i^{(r)}, D_i^{(t)})$$

$$= \frac{1}{K} \sum_{C_i^{(r)}} p(D_i^{(t)}|C_i^{(r)}, C_i^{(t)}, I_i^{(r)}) p(C_i^{(t)}|C_i^{(r)}) p(C_i^{(r)}|I_i^{(r)}), \quad (1)$$

where we have assumed that $C_i^{(t)}$ is conditionally independent of $I_i^{(r)}$ given $C_i^{(r)}$. The right side of equation (1) can be seen as a product of three terms: a



(a) $\mathcal{N}_i^k$ for k=0-3 (b) $\{C_{\mathcal{N}_i}^{(t)}\}$

Fig. 2. (a) shows $\mathcal{N}_i^k$ for k=0-3 while (b) shows a sample tissue label configuration, where $\mathcal{N}_i$ is defined as the 4-voxel neighbourhood in both (a) and (b)

difference likelihood, a tissue transition likelihood, and a prior classification of our reference image. We incorporate information from neighbouring voxels using a neighbourhood likelihood model and by recursively growing our observation space, $\mathcal{N}_i^k$, defined as

$$\mathcal{N}_i^k = \bigcup_{j \in \mathcal{N}_i^{k-1}} \mathcal{N}_j, \tag{2}$$

where $\mathcal{N}_i^0$ is just the voxel in question (as in (1)), and $\mathcal{N}_i$ is a the first-order neighbourhood. We can express the posterior probability of a class tissue, $C_i^{(t)}$ for a k^{th} inference problem as

$$p_i^k = p(C_i^{(t)}|I_{\mathcal{N}_i^k}^{(r)}, D_{\mathcal{N}_i^k}^{(t)}) = \sum_{\{C_{\mathcal{N}_i}^{(t)}\}} p(C_i^{(t)}, \{C_{\mathcal{N}_i}^{(t)}\}|I_{\mathcal{N}_i^k}^{(r)}, D_{\mathcal{N}_i^k}^{(t)}) \tag{3}$$

$$= \sum_{\{C_{\mathcal{N}_i}^{(t)}\}} p(\{C_{\mathcal{N}_i}^{(t)}\}|I_{\mathcal{N}_i^k}^{(r)}, D_{\mathcal{N}_i^k}^{(t)}) p(C_i^{(t)}|\{C_{\mathcal{N}_i}^{(t)}\}, I_i^{(r)}, D_i^{(t)})$$

$$= \sum_{\{C_{\mathcal{N}_i}^{(t)}\}} \frac{P^k_{\{C_{\mathcal{N}_i}^{(t)}\}}}{K^k} \sum_{C_i^{(r)}} p(D_i^{(t)}|C_i^{(r)}, C_i^{(t)}, I_i^{(r)}) p(\{C_{\mathcal{N}_i}^{(t)}\}|C_i^{(t)}) p(C_i^{(t)}|C_i^{(r)}) p(C_i^{(r)}|I_i^{(r)}), \tag{4}$$

where K^k is a normalization constant, $P^k_{\{C_{\mathcal{N}_i}^{(t)}\}}$ is the probability of a tissue class label configuration, $\{C_{\mathcal{N}_i}^{(t)}\}$ is a configuration of tissue class labels in $\mathcal{N}_i$ (see example in Fig. 2b) and where the summation implies that we consider all possible configurations. We assume Markovianity ($p(C_i^{(t)}|C_{j \notin \mathcal{N}_i}, C_{\mathcal{N}_i}^{(t)}) = p(C_i^{(t)}|C_{\mathcal{N}_i}^{(t)})$), causality ($p(C_i^{(r)}|C_{\mathcal{N}_i}^{(t)}) = p(C_i^{(r)})$), conditional independence of $D_i^{(t)}$ from $C_{\mathcal{N}_i}^{(t)}$ given $C_i^{(t)}$, and conditional independence of $C_{\mathcal{N}_i}^{(t)}$ from $C_i^{(r)}$ given $C_i^{(t)}$. We can expand $P_{\{C_{\mathcal{N}_i}^{(t)}\}}$ as follows:

$$P^k_{\{C_{\mathcal{N}_i}^{(t)}\}} = p(\{C_{\mathcal{N}_i}^{(t)}\}|I_{\mathcal{N}_i^k}^{(r)}, D_{\mathcal{N}_i^k}^{(t)}) = \prod_{j \in \mathcal{N}_i} p(C_j^{(t)}|I_{\mathcal{N}_j^{k-1}}^{(r)}, D_{\mathcal{N}_j^{k-1}}^{(t)}) = \prod_{j \in \mathcal{N}_i} p_j^{k-1}. \tag{5}$$

This iterative process can be seen as modelling tissue label dependencies locally while recursively growing the observation space, which, in the limit, would consider the entire image.

Likelihood Models

Difference likelihood models are learned from training data for all combinations of $C^{(r)}$ and $C^{(t)}$, where tissue classes are restricted to one of 5 classes (csf, gm, wm, les, pv). Transitions from $C^{(r)}$ to $C^{(t)}$ may represent real change (e.g. wm-les), misregistration (e.g. wm-gm), differences in partial volume effects at different timepoints, variability or error in segmentation of the training data, or some combination of these factors. To reduce the number of models that need to be learned, $D^{(t)}$ is assumed to be conditionally independent of $I^{(r)}$ given $C^{(r)}$, except for the case of $C^{(r)}$=les. For this special case, we use PCA to project our multimodal intensities onto a 1-D subspace that best captures intensity variations seen in lesions, and separate this subspace into 3 distinct lesion intensity classes. Difference likelihood models for wm-wm, gm-gm, csf-csf and pv-pv transitions are modeled as 3D Gaussians as they are well approximated as such. All other models are represented by 3D non-parametric distributions using Parzen windows [11].

The neighbourhood likelihood represents the likelihood of observing a neighbourhood configuration $C^{(t)}_{\mathcal{N}_i}$ around a voxel with label $C^{(t)}_i$. A 4-voxel in-plane neighbourhood was used, and a neighbourhood configuration was represented by a count of each tissue class in the 4-voxel neighbourhood. Models were constructed as histograms by observing the frequencies of the different neighbourhood representations that occured in training data for each tissue class.

The transition likelihood represents the prior probability of transitioning from one tissue label to any other (or the same) tissue label. This acts as a bias toward the tissue class at the reference timepoint and towards more plausible tissue label transitions. Transition likelihoods are learned based on frequency of occurence in the training data.

The inclusion of a prior term, $p(C^{(r)}|I^{(r)})$, implies that we have available some form of probabilistic tissue classification for the reference timepoint. This prior on $C^{(r)}$ can come from some other automated tissue segmentation scheme, from manual labeling, or from some combination of both.

3 Experiments

3.1 Preprocessing

All MRI data used in this study consists of sets of T1-weighted (T1), T2-weighted (T2) and PD-weighted (PD) images at a resolution of 1x1x3mm. Each scan was corrected for intensity non-uniformity [12], masked to exclude non-brain and the posterior fossa, linearly (6 DOF) registered to the T2 image at the baseline scan from the same patient, and intensity normalized to a common intensity space [13]. The "ground truth" tissue labels were generated from an automatic tissue segmentation using an in-house classifier based on [14] which then had voxels classified as lesion manually verified and corrected by experts. Tissue class labels were restricted to one of csf, gm, wm, les, and pv. These manually corrected (MC) 5-tissue class segmentations served as a reference for subsequent training and validation.

3.2 Scan-Rescan

A scan-rescan data set consisting of 3 relapsing-remitting MS (RRMS) patients allowed us to validate the precision of the proposed method in the absence of real physical change. Patients were scanned on a Siemens Sonata 1.5T scanner, removed from the scanner, and then rescanned. For the scan-rescan data set, fully manual lesion labels (FM) were also available in addition to the manually corrected lesion labels (MC). Lesions were required to consist of at least two contiguous voxels. The MC labels for the reference timepoint were used as a prior for the Bayesian classifier (BC). Classification was done by first using the scan as the reference timepoint (BC-S) and classifying the rescan, and secondly using the rescan as the reference (BC-R) and classifying the scan. Models used for classification were learned from independent training data. We define new lesion voxels as those that were not labelled as lesion in the reference scan but labelled as lesion in the follow-up scan, and resolved lesions those that were labelled as lesion at reference but not in the follow-up. Means and standard deviations of lesion volume at reference, new lesion voxels, resolved lesion voxels and change in lesion volume over the 3 scan-rescan patients are shown in Table 1. Given that there is no biological change in the scan-rescan period, ideally no lesion activity would be detected. The number of new and resolving lesion voxels are greatly reduced when using the proposed Bayesian classifier as compared to both the FM and MC labels, suggesting greater precision with the proposed method.

Table 1. Scan-Rescan precision for 3 RRMS patients

	MC	FM	BC-S	BC-R
Lesion Volume at Reference (voxels)	7466±4278	7517±4098	7466±4278	7466±4404
New Lesion Voxels	1368±853	1657±887	74±8	26±9
Resolved Lesion Voxels	1313±700	1453±906	49±21	28±15
Net Change in Lesion Voxels	55±155	204±321	25±27	-3±20

3.3 Clinical Data

Increased precision is only meaningful if the classifier is still sensitive to true change. A clinical data set was used to validate the sensitivity of the proposed method to new and enlarging lesions. This data set consists of 212 total scans from 89 RRMS patients with 2-4 longitudinal scans each, taken over a period of 48 weeks with a minimum interval of 12 weeks between scans. Fully manual lesion labels were not available for this data set, so MC labels were used as reference for all 212 scans. Meaningful evaluation based on comparison to reference lesion labels is challenging, due to lack of consensus as to what consitutes a lesion, ambiguity of lesion boundaries, and lack of precision in labelling of the same patient over time. A subset of the new lesion voxels from the MC labels were identified as being new lesions or enlarging portions of existing lesions, based on a minimum of 3 contiguous new lesion voxels and spatial properties

Table 2. Apparent Sensitivity to New and Enlarging Lesions as compared to MC Labels

NE Size	ALL	>=5 voxels	>=10 voxels
Total # NE	63	58	45
Criteria(a)	46 (73%)	45 (78%)	37 (82%)
Criteria(b)	53 (84%)	52 (90%)	44 (98%)
Voxel-wise sensitivity	76.6%	76.9%	77.5%

of connectedness to existing lesions. This set of new or enlarging (NE) lesion labels was manually verified by experts and ensured as much as possible that our ground truth definition of NE lesions corresponds to real change in brain tissue. For each timepoint other than the baseline scan, the scan and MC tissue labels from the previous timepoint were used as the reference image and prior. In this way, all scans except for the baseline scan were classified in a pairwise fashion. Four-fold cross-validation was used, with 66 or 67 patients used for training our models, and 22 or 23 used for testing, on each fold. Performance of our classifier was measured based on the number of NE lesions that were detected. Two different criteria were used to decide whether an NE lesion was considered as detected : (a) identification of a minimum of 50% of voxels in an NE lesion and (b) identification of 3 or more voxels in an NE. Analysis was done separately for all NE lesions and subsets of NE lesions that were greater than 5 and 10 voxels in size. A voxel-wise sensitivity to NE lesions was also measured, which is defined as the percentage of all voxels in new and enlarging lesions that were classified as lesion.

Specificity of newly detected lesions was not quantitatively evaluated as we did not have a filtered subset of MC labels that allowed for a meaningful comparison. Qualitative analysis showed that for a small subset of scans, significant false detection of new lesions occured adjacent to the lateral ventricles. Distortion, atrophy and partial volume effects in the z-direction all contributed to these false detections. Sensitivity and specificity of resolving lesions were also not explicitly measured due to lack of suitable ground truth. Qualitative analysis showed good sensitivity to fully resolving lesions, but resolving portions of partially resolving lesions were generally underestimated, due to the attractive effect of remaining lesion in the neighbourhood likelihood model. Very few falsely resolving lesions were observed.

Table 3. Number of New And Resolving Lesion Voxels for Slices in Figure 3

	Slice 1		Slice 2		Slice 3	
	MC	BC	MC	BC	MC	BC
Lesion Voxels At Reference	156	156	181	181	347	347
New Lesion Voxels	58	57	100	42	137	2
Resolved Lesion Voxels	77	30	77	41	108	3
Net Change in Lesion Voxels	-19	27	33	1	21	-1

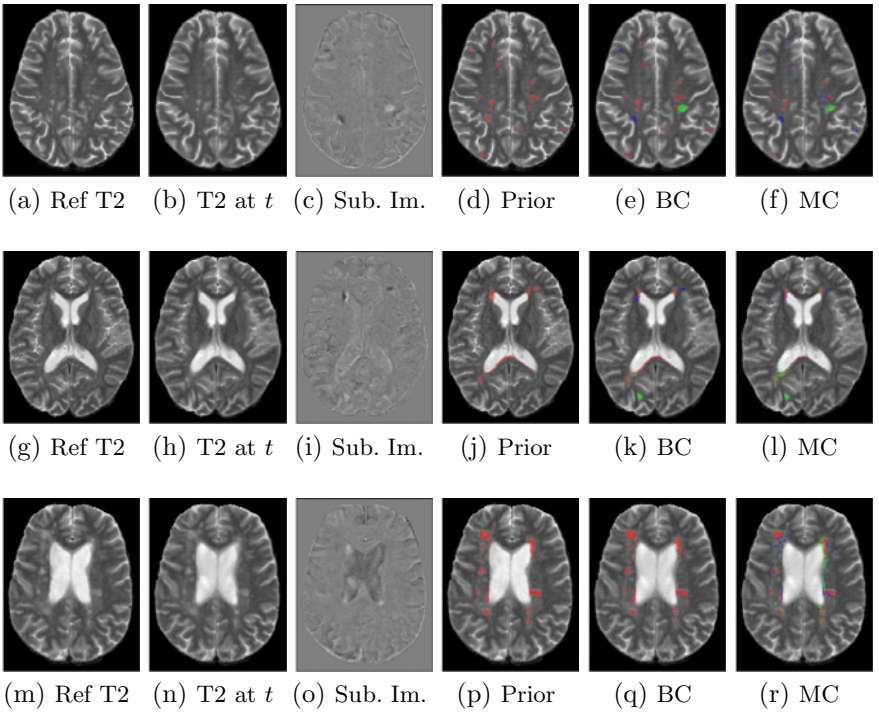


Fig. 3. Sample classification results for three slices from three different patients, where Slice 1 (a-f) and Slice 2 (g-l) both have new and resolving lesions and Slice 3 (m-r) exhibits little or no real change. Reference T2 images (Ref. T2), T2 images at the timepoint t to be classified (T2 at t), subtraction images between the two timepoints (Sub. Im.), prior classifications at the reference timepoints (Prior), output of the proposed Bayesian classifier for time t (BC), and MC labels for time t (MC), are shown for each slice. The BC and MC labels are colour-coded as follows: stable lesion voxels are shown in red, new lesion voxels are shown in green, and voxels that were lesion at the reference timepoint but have resolved are shown in blue.

Sample classification results for 3 slices of 3 different patients in the clinical data set are shown in Figure 3, and the lesion activity in those slices as detected by the proposed classifier and the MC labels is summarized in Table 3. New and resolving lesions are correctly identified while stable lesions are labelled in a much more consistent manner than for the MC labels, where lesion boundaries are shown to fluctuate, and more ambiguous tissue intensities may be labeled differently at the two timepoints despite lack of apparent change.

4 Discussion and Future Work

In this paper, we introduce an automatic Bayesian classifier that detects MS lesion activity in longitudinal scans based on subtraction images. Our approach

attempts to overcome the limitations of subtraction images in terms of registration error and noise by embedding a prior classification at a reference timepoint and by building likelihood models that account for artifact and noise. Our approach was evaluated on both a scan-rescan data set and a large multicenter clinical data set and has demonstrated increased precision as compared to a manual classification, while remaining sensitive to lesion activity.

A quantitative evaluation of sensitivity to resolving lesions and specificity of both new and resolving lesions is needed to fully characterize the performance of the proposed classifier. The incorporation of non-linear registration or explicit segmentation of lateral ventricles may aid in reducing false detection of new lesions in patients where there is significant atrophy or distortion. The preprocessing pipeline used was chosen based on convenience. More optimal pipelines specific to longitudinal data may help reduce noise and artifact in subtraction images [9,15].

References

1. Lee, M.A., Smith, S., et al.: Defining multiple sclerosis disease activity using MRI T2-weighted difference imaging. Brain 121, 2095–2102 (1998)
2. Tan, I.L., van Schijndel, R.A., et al.: Image Registration and subtraction to detect active T_2 lesions in MS: an interobserver study. J. Neurol. 249, 767–773 (2002)
3. Moraal, B., Meier, D.S., et al.: Subtraction MR Images in a Multiple Sclerosis Multicenter Clinical Trial Setting. Radiology 250, 506–514 (2009)
4. Duan, Y., Hildenbrand, P.G., et al.: Segmentation of Subtraction Images for the Measurement of Lesion Change in Multiple Sclerosis. Am. J. Neuroradiol. 29, 340–346 (2008)
5. Rey, D., Subsol, G., et al.: Automatic detection and segmentation of evolving processes in 3D medical images: Application to multiple sclerosis. Med. Image Anal. 6, 163–179 (2002)
6. Welti, D., Gerig, G., et al.: Spatio-temporal Segmentation of Active Multiple Sclerosis Lesions in Serial MRI Data. In: Insana, M.F., Leahy, R.M. (eds.) IPMI 2001. LNCS, vol. 2082, p. 438. Springer, Heidelberg (2001)
7. Prima, S., Arnold, D.L., et al.: Multivariate Statistics for Detection of MS Activity in Serial Multimodal MR Images. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 663–670. Springer, Heidelberg (2003)
8. Aït-Ali, L.S., Prima, S., et al.: STREM: A Robust Multidimensional Parametric Method to Segment MS Lesions in MRI. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 409–416. Springer, Heidelberg (2005)
9. Bosc, M., Heitz, F., et al.: Automatic change detection in mutimodal serial MRI: application to multiple sclerosis lesion evolution. NeuroImage 20, 643–656 (2003)
10. Thirion, J.-P., Calmon, G.: Deformation Analysis to Detect and Quantify Active Lesions in Three-Dimensional Medical Image Sequences. TMI 18, 429–441 (1999)
11. Turlach, B.: Bandwidth selection in kernel density estimation: a review. Discussion paper 9317, Institut de Statistique, UCL, Louvain la Neuve, Belgium (1993)
12. Sled, J.G., Zijdenbos, et. al.: A non-parametric method for automatic correction of intensity nonuniformity in MRI data. TMI 17, 87–97 (1998)
13. Nyùl, L.G., Udupa, J.K., et al.: New variants of a method of MRI scale standardization. TMI 19, 143–150 (2000)
14. Francis, S.: Automatic lesion identification in MRI of MS patients. Master's Thesis, McGill University (2004)
15. Meier, D.S., Guttman, R.G.: Time-series analysis of MRI intensity patterns in multiple sclerosis. NeuroImage 20, 1193–1209 (2003)

Multivariate Network-Level Approach to Detect Interactions between Large-Scale Functional Systems

Wei Gao[1,4], Hongtu Zhu[2,4], Kelly Giovanello[3,4], and Weili Lin[1,4]

Department of [1] Radiology, [2] Biostatistics and [3] Psychology and [4] Biomedical Research Imaging Center
University of North Carolina at Chapel Hill, NC 27599, USA
wgao@email.unc.edu

Abstract. The question of how large-scale systems interact with each other is intriguing given the increasingly established network structures of whole brain organization. Commonly used regional interaction approaches, however, cannot address this question. In this paper, we proposed a multivariate network-level framework to directly quantify the interaction pattern between large-scale functional systems. The proposed framework was tested on three different brain states, including resting, finger tapping and movie watching using functional connectivity MRI. The interaction patterns among five predefined networks including dorsal attention (DA), default (DF), frontal-parietal control (FPC), motor-sensory (MS) and visual (V) were delineated during each state. Results show dramatic and expected network-level correlation changes across different states underscoring the importance of network-level interactions for successful transition between different states. In addition, our analysis provides preliminary evidence of the potential regulating role of FPC on the two opposing systems-DA and DF on the network level.

1 Introduction

Recent development of functional connectivity magnetic resonance imaging (fcMRI) [1] has greatly improved our understanding of the brain's functional organization. Multiple functional networks including motor-sensory (MS), visual (V), dorsal attention (DA), and more recently the default (DF) networks have been delineated using this technique [1, 2]. These findings have greatly improved our understanding of whole brain functional segregation and integration and reinforced the notion that normal brain functioning relies on coordinated activity among sets of distributed yet interacted brain regional/functional systems.

Yet, how different networks interact with each other remains elusive. Fox et al [3] looked into the interaction between the DF and DA networks and observed the existence of anti-correlation between them even during a resting state. Kelly et al [4] further revealed that the strength of this negative correlation is modulated by exogenous demands and is positively correlated with behavioral performance, underscoring the functional significance of between-network interactions. Moreover, other groups [5] have recently proposed the "network degeneration hypothesis" where they reported that neurodegeneration may be related to network-level dysfunction and

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 298–305, 2010.

suggested the need of developing new network-based diagnostic assays. Therefore, approaches to specifically reveal brain network-level interactions may have profound implications not only in informing normal brain function but also offering means to discern brain network dysfunction in patients with neurodegenerative diseases.

While early studies provide invaluable insights into network interaction, one potential limitation is that a correlation between two mean time courses which were derived through averaging time signals over all brain regions within a network was commonly employed to quantify network level interaction [4, 6]. This essentially ignores the inter-dependence structure within a network. Although facilitating theoretical interpretation and utilizing a straightforward computation, this simplification is prone to information loss and/or even becomes untenable when the homogeneity assumption is violated.

In this study, we aimed to develop a multivariate network-level framework capable of discerning functional network interactions by integrating several statistical tools, including canonical correlation analysis (CCA) [7], multivariate regression, and a resampling method. Compared with the existing methods, three novel features make our method suitable for network-level analysis: (i) by applying multivariate techniques, our method is free from information loss induced by averaging; (ii) through dimension normalization, our method is able to handle interactions between networks of arbitrary sizes and avoid potential biases; and (iii) leveraging multiple regression, our method is not only able to study network-level correlation but also partial correlation, adding another dimension to functional analysis.

Five networks, including the mentioned MS, V, DA, DF, and another more recently delineated frontal parietal control (FPC) network [8] were included to test the proposed method. Besides, three experimental states, including resting, finger tapping and movie watching were examined to (i) detect the dynamics of network-level correlation and (ii) test the hypothesis of the FPC's regulation role during different cognitive states. Results showed extensive network-level correlation changes for successful transition between different states and provided preliminary evidence of FPC's regulating role on the two opposing systems, the DA and DF.

2 Methods

We develop a multivariate network-level approach for directly quantifying the correlation/partial correaltion pattern between large-scale functional systems. The overall flow chart of the proposed method is illustrated in Figure 1.

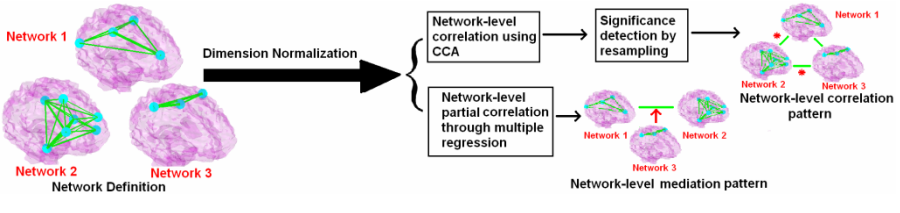


Fig. 1. Flow chart of the multivariate network-level framework

2.1 Multivariate Network Correlation Using Canonical Correlation Analysis

Time courses from individual regions within each network were averaged first to represent the temporal characteristic of the whole network. Given the obvious dimension differences among different networks, there is a potential bias in the network-level interaction calculation. To overcome this bias, principle component analysis (PCA) was performed and the top 4 principle components for each network were selected for subsequent multivariate calculations. A retrospective check reveals that more than 98% percent of the variance is preserved for all networks, but the number of components can be adjusted in a study-dependent manner.

To quantify the network level interaction, a canonical correlation measure between two multivariate vectors was used. Canonical correlation analysis [7] has been widely employed to measure the association between multivariate variables: $X = [x_1,..., x_m]^T$ and $Y = [y_1,..., y_n]^T$. Particularly, for two univariate random variables, the canonical correlation is identical to the Pearson's correlation. The key idea of CCA is to maximize the correlation between linear combinations of X and Y denoted by $U_1 = w_1 x_1 + ... + w_m x_m = w^T x$ and $U_2 = v_1 y_1 + ... + v_n y_n = v^T y$, respectively. The canonical correlation equals to the largest eigenvalue of the matrix of $C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx}$ (or $C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy}$). Notice the value of network canonical correlation (NCC_{xy}) is between 0 and 1 with 0 indicating no dependence and 1 indicating full dependence.

2.2 Network-Level Partial Correlation through Multivariate Regression

In bivariate statistics, partial correlation is used to assess potential mediation effect, which is defined as the correlation between two random variables after controlling for another "mediator" variable. Specifically, partial correlation $\rho_{xy/z}$ between two random variables x and y controlling for another variable z can be computed as the Pearson's correlation between the residuals ε_x and ε_y from two linear regression equations:

$$x = x_0 + \beta_1 z + \varepsilon_x \tag{1}$$

$$y = y_0 + \beta_2 z + \varepsilon_y \tag{2}$$

For mediation analysis, $\rho_{xy/z}$ denotes the correlation remained between x and y when all mediating effect of z is removed and the differences between the ordinary correlation ρ_{xy} and $\rho_{xy/z}$, represent the amount of mediation that z exerts on the relation between x and y.

Since the primary focus of this study is to depict correlation between two sets of variables, the above computational procedures are generalized to accommodate the multivariate property of this problem. For two sets of multivariate vectors $X = [x_1,..., x_m]^T$ and $Y = [y_1,..., y_n]^T$, and another set of independent

covariates $Z = [z_1,..., z_p]^T$, influences from which will be removed, Eq.1 and Eq. 2 can be written as:

$$X = X_0 + B_1 Z + E_x \quad (3)$$

$$Y = Y_0 + B_2 Z + E_y \quad (4)$$

where E_x and E_y are the residual vectors of X and Y, respectively, after regressing on the variable set of Z. Subsequently, the canonical correlation coefficient of E_x and E_y can be calculated, representing the partial correlation $NPC_{XY/Z}$ between the two sets of variables (networks) X and Y. The difference of $NCC_{XY} - NPC_{XY/Z}$ thus represents the potential mediation effects of Z exerting on the relation between X and Y.

2.3 Resampling to Define Significance of Network-Level Correlation

After calculating network-level correlations, the next question is to determine whether a certain correlation is statistically significant. Rather than testing the null hypothesis of zero interaction using random signal, we prompt to use a null distribution derived from resampling of the actual brain signals. Specifically, this null distribution is obtained by randomly permuting a set of reference regions into two sets and calculating their network level correlations. The reference regions are chosen to be those that demonstrate the minimum possible interactions (according to the absolute values of Pearson's correlation) among each other from the 90 pre-defined brain regions covering the whole brain [9]. The rationales behind this choice are (i) the wide spread dependence (either positive or negative correlation) between different brain regions and (ii) the large number of regions involved in each network. This means that even by randomly selecting a number of regions (among the least interacted regions) to construct two networks, certain amount of dependence between them is possible, which can be reasonably defined as the "baseline" for network-level correlations. As a result, network-level interactions that are comparable to this baseline should be excluded and only those that are statistically higher should be considered.

More specifically, based on the 90 ROI template [9], we selected a set of 15 reference regions during each state, which were then randomly distributed to 2 networks with 6 and 9 regions (equal to the largest possible combination of number of regions in the five selected networks, details in Section 3) and the network-level correlations were calculated 1000 times to generate the null distribution for each state, from which the p-value for each correlation can be defined using appearance ratio. This process was done for each subject and the obtained p-values were combined across subjects to give a group p-value using the Fisher's method [10] for each correlation. Significant correlations were defined at $\alpha = 0.05$ after correcting for multiple comparisons using FDR [11]. The same procedures were done for each of the three states examined.

3 Materials

A total of 19 healthy subjects (age 25~33, 7F, all right-handed) were recruited in this study. A T2*-weighted EPI sequence was used to acquire fcMRI signal on a 3T scanner: TR = 2sec, TE = 32 ms; 33 slices; and voxel size = 4x4x4 mm^3. This sequence was repeated 150 times (~5 min) for each of the three states. During resting, subjects were instructed to relax and remain still but keep eyes closed. During finger tapping, subjects were instructed to continuously touching the thumb to each of the other fingers in a sequential manner at a roughly consistent pace (~1Hz) with eyes closed. For movie watching, the movie clip contains shallow sea scenes.

After standard preprocessing steps, including time shifting, motion correction, spatial smoothing (6-mm FWHM Gaussian kernel), and band filtering (<0.08Hz), nuisance sources of variance (white matter, CSF and the mean global signal) were removed through regression. Three subjects were excluded from the subsequent analysis because of excessive head motion. Images of the first ten time points were excluded to allow magnetization reaching a steady state. Spatial normalization to MNI template was achieved using nonlinear registration and regional parcellation was done base on Tzourio-Mazoyer et al [9].

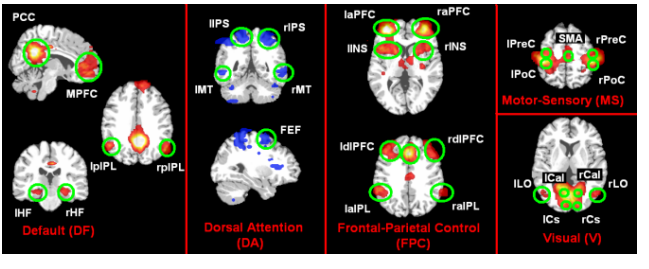


Fig. 2. ROIs defined for the five networks based on functional correlation map

The network definition is similar as that done by Van Dijk et al [2], which is based on the combined consideration of (i) prior knowledge of network composition and (ii) the study-specific functional correlation map. Specifically, regions of interest (ROIs) were defined as 8-mm spheres around the detected peaks (high correlation values) to construct the corresponding networks. For peak definition, an initial seed in the posterior cingulate cortex (PCC) was used to compute a group mean correlation map which was then used to indentify peaks within the DF network, including the medial prefrontal cortex (MPFC), bilateral inferior parietal lobule (IPL) and hippocampus formation (HF). Given the well-established anti-correlation with DF [3], the negative peaks of the PCC correlation map were identified for regions in the DA network, including the bilateral intra-parietal sulcus (IPS), frontal eye field (FEF) and middle temporal area (MT+). For FPC, the seeds were centered at the bilateral aPFC [8] to define peaks in nodes, including the anterior cingualte cortex (ACC), dorsolateral prefrontal cortex (dlPFC), anterior insula (aINS), and anterior inferior parietal lobule (aIPL). In addition, the MS and V networks were similarly defined using the bilateral postcentral and calcarine cortex as initial seeds, respectively, to define peaks in

precentral/postcentral gyrus (PreC, PoC), supplementary motor-sensory area (SMA) for MS and bilateral calcarine (Cal), cuneus (CS), and lateral occipital (LO) for V. Note all functional network definitions were based on the resting state data and altogether there are 6, 9, 6, 5, and 6 regions for DA, FPC, DF, MS and V, respectively, which are presented in Fig.2.

4 Experimental Results

The network-level correlation patterns using the spring-embedding plots are presented in Fig.3. The width of edges corresponds to the strength of interactions. Moreover, the statistical grouping of the set of interaction values are also achieved using the Tukey's test at $\alpha = 0.05$ to detect those that are significantly higher than the rest during each state (red asterisks in Fig.3).

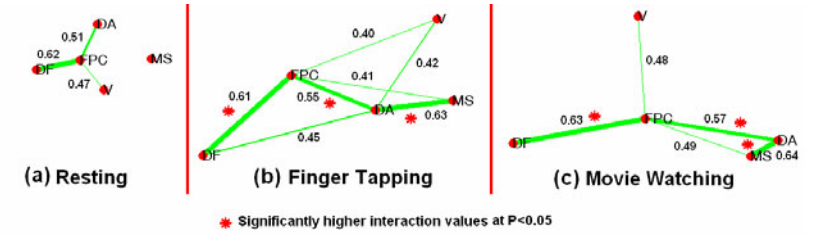


Fig. 3. The network-level correlation pattern during the three examined states. The value besides each connection represents the group average network-level correlation. Red asterisks indicate statistically stronger correlations using Tukey's test at $\alpha = 0.05$.

During the resting state, only three significant connections exist with FPC at the center and connected with DA, DF, and V while MS is left alone. This sparse structure shows minimal interactions among networks during resting but the center position of FPC implies its significant "bridging" role, which will be specifically tested later. During finger tapping, the interaction pattern becomes much more extensive: besides the observed 3 interactions during resting, DA now significantly interacts with MS, which is highly expected given the apparent attentional requirement. In addition, this observation could also suggest DA's top-down control over MS for successful task performance[12]. Moreover, the DA-DF also becomes significant which is consistent with the previously reported increased anti-correlation between them during goal-directed tasks [4]. Other significant interactions include DA-V, and FPC-MS. During movie watching, the changes are less extensive but still two more interactions (DA-MS, FPC-MS) emerge which might be due to the increased control over eye movement. Overall, minimal network-level interactions during resting together with the enhanced interactions during the two task states indicate increased coordination/competition between networks during task performance.

Another interesting pattern is that interactions between FPC-DA and FPC-DF persist across all three states and they are statistically stronger than other interactions

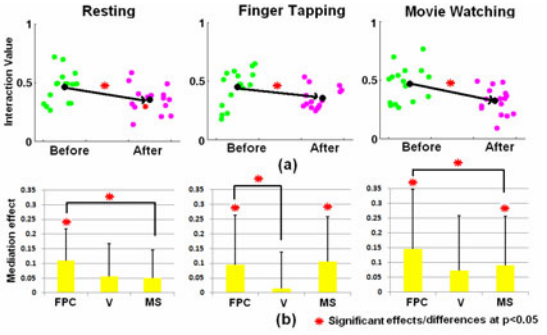


Fig. 4. Mediation analysis on the network-level interaction between DA and DF during three examined states. (a) mediation effect of FPC by comparing the interaction values (DA-DF) before (green dots) and after regressing out its effect (pink dots); (b) comparisons of FPC's mediation effect on DA-DF with that of V and MS across all three states.

during both finger taping and movie watching (except DA-MS, which is comparable, Fig.3). These findings clearly indicate the importance of this triad system and are consistent with the hypothesis [8] that FPC might regulate the two opposing systems- DA and DF. To directly test this hypothesis, network-level partial correlation between DA and DF using FPC as a control network was calculated during each state and the interaction values before and after this regression were compared to test the mediation role of FPC between these two systems. The results are shown in Fig.4a, where a significant reduction of interaction strength is observed for all three states ($p=0.0012$, 0.0377, and 0.0105 using one way ANOVA), strongly supporting FPC's regulation role on DA and DF. Moreover, this effect (by taking the difference between values before and after regressing out of its effect) is compared among FPC, MS, and V (Fig.4b), where it shows that (i) FPC is the only network demonstrating significant correlation-reduction effect during resting; (ii) V fails to show any significant effect during any state; and (iii) although MS shows significant effect during the two task states, its effect is significantly weaker than that of FPC during movie watching. Note although causal relationship should be validated before concluding this regulating role, our results provide initial support for FPC's regulating role over the two most salient, anti-correlated systems- DA and DF on the network level.

5 Conclusion

In this paper, we proposed a new multivariate network-level framework to quantify the correlations/partial correlations between large-scale systems and demonstrated its application during three different brain states, including resting, finger tapping and movie watching. Through direct application of multivariate techniques, our method is (i) free of information loss induced by averaging; (ii) able to handle networks of arbitrary size through dimension normalization; and (iii) able to detect potential mediation effects through multiple regressions. Experimental results show dramatic and extensive network correlation changes which are consistent with our expectations

and previous findings [8], underscoring the importance of network level coordination in task fulfillment. Moreover, network-level partial correlation analysis provides support for the potential regulating role of FPC over the two opposing systems-DA and DF.

Acknowledgement. This work was supported in part by NSF (BCS-08-26844) and NIH (NS RO1055754, UL1-RR025747-01, MH086633, P01CA142538-01 and AG033387).

References

1. Biswal, B.B., Yetkin, F.Z., Haughton, V.M., Hyde, J.S.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magn. Reson. Med. 34, 537–541 (1995)
2. Van Dijk, K.R., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., Buckner, R.L.: Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. J. Neurophysiol. 103, 297–321 (2010)
3. Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E.: The human brain is intrinsically organized into dynamic, anticorrelated functional networks. Proc. Natl. Acad. Sci. USA 102, 9673–9678 (2005)
4. Kelly, A.M., Uddin, L.Q., Biswal, B.B., Castellanos, F.X., Milham, M.P.: Competition between functional brain networks mediates behavioral variability. Neuroimage 39, 527–537 (2008)
5. Seeley, W.W., Crawford, R.K., Zhou, J., Miller, B.L., Greicius, M.D.: Neurodegenerative diseases target large-scale human brain networks. Neuron. 62, 42–52 (2009)
6. Jafri, M.J., Pearlson, G.D., Stevens, M., Calhoun, V.D.: A method for functional network connectivity among spatially independent resting-state components in schizophrenia. Neuroimage 39, 1666–1681 (2008)
7. Hotelling, H.: Relations between two sets of variants. Biometrika, 312–377 (1936)
8. Vincent, J.L., Kahn, I., Snyder, A.Z., Raichle, M.E., Buckner, R.L.: Evidence for a frontoparietal control system revealed by intrinsic functional connectivity. J. Neurophysiol. 100(6), 3328–3342 (2008)
9. Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M.: Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage 15(1), 273–289 (2002)
10. Lazar, N.A., Luna, B., Sweeney, J.A., Eddy, W.F.: Combining brains: a survey of methods for statistical pooling of information. Neuroimage 16(2), 538–550 (2002)
11. Benjamini, Y., Yekutieli, D.: The contorl of the false discovery rate in multilpe testing under dependency. Ann. Statist. 29, 1165–1188 (2001)
12. Gordon, A.M., Lee, J.H., Flament, D., Ugurbil, K., Ebner, T.J.: Functional magnetic resonance imaging of motor, sensory, and posterior parietal cortical areas during performance of sequential typing movements. Exp. Brain Res. 121(2), 153–166 (1998)

A Generalized Learning Based Framework for Fast Brain Image Registration

Minjeong Kim, Guorong Wu, Pew-Thian Yap, and Dinggang Shen

Department of Radiology and BRIC, University of North Carolina at Chapel Hill
{mjkim,grwu,ptyap,dgshen}@med.unc.edu

Abstract. This paper presents a generalized learning based framework for improving both speed and accuracy of the existing deformable registration method. The key of our framework involves the utilization of a support vector regression (SVR) to learn the correlation between brain image appearances and their corresponding shape deformations to a template, for helping significantly cut down the computation cost and improve the robustness to local minima by using the learned correlation to instantly predict a good subject-specific deformation initialization for any given subject under registration. Our framework consists of three major parts: 1) *training* of SVR models based on the statistics of image samples and their shape deformations to capture intrinsic image-deformation correlations, 2) *deformation prediction* for a new subject with the trained SVR models to generate a subject-resemblance intermediate template by warping the original template with the predicted deformations, and 3) *estimating of the residual deformation* from the intermediate template to the subject for refined registration. Any existing deformable registration methods can be easily employed for *training* the SVR models and *estimating the residual deformation*. We have tested in this paper the two widely used deformable registration algorithms, i.e., HAMMER [1] and diffeomorphic demons [2], for demonstration of our proposed frameowrk. Experimental results show that, compared to the registration using the original methods (with no deformation prediction), our framework achieves a significant speedup (6X faster than HAMMER, and 3X faster than diffeomorphic demons), while maintaining comparable (or even slighly better) registration accuracy.

1 Introduction

Deformable registration of brain images has been extensively applied for facilitating identification of brain abnormality by comparison between normal and abnormal groups, and for constructing atlases to reflect structural and functional variation of individuals within a population. Most deformable registration methods, regardless intensity- or feature-based, aim to estimate deformation fields for better establishing structural correspondences between images. Although many methods have been previously proposed, determining a reasonable template-subject deformation remains a challenge due to large inter-subject structural variations.

It is worth noting that the traditional registration methods usually lack a good deformation field initialization mechanism. This generally results in long computation

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 306–314, 2010.

time arising from the need of estimating large complex deformations, and also vulnerability to misleading matching due to structural ambiguities. To attack these problems, we propose a learning based method to improve registration speed and accuracy by predicting a good initial deformation for bringing the template closer to the given subject. In this way, the remaining deformation from the *intermediate template* (IT) to the given subject becomes small, and thus can be easily estimated by many deformable registration methods.

Recently, learning-based statistical deformation models [3-5] based on deformation features, i.e., wavelet coefficients, B-spline coefficients, or principal components, have been employed to improve registration accuracy by imposing more realistic registration constraints. Although these approaches have greatly improved registration accuracy, computation cost required for constraining and estimating the deformation field is still quite high.

In this paper, we propose a generalized learning based framework for fast deformable registration, the key of which involves employing support vector regression (SVR) models for rapidly generating intermediate templates. Specifically, we first simulate a large number of training samples with the built image appearance and deformation statistical models. Next, we learn the SVR models for correlating image appearance features and their deformation coefficients (to the template). The learnt SVR models are finally applied to rapidly predict a good initial deformation for a given new subject.

To our knowledge, this current work presents the first attempt to combine, via SVR, the statistics of brain image appearances and their deformation coefficients for effectively guiding deformable registration. It is not difficult to see that many conventional registration algorithms can be easily integrated into our framework for immediate improvement, since our framework is based upon image appearances and deformation coefficients, not on a specific deformation model.

Experimental results on real brain images indicate that this SVR-based deformation prediction approach allows multi-fold speedup of deformable registration algorithms, i.e., 6X and 3X faster for HAMMER and diffeomorphic demons, respectively, while retaining similar (or even slightly better) registration accuracy.

2 Method

The goal of a deformable registration algorithm is to estimate a dense transformation field u for aligning a subject image S to a template T, as shown in **Fig. 1**. Our approach involves decomposing the overall deformation field u into two parts: the estimated initial deformation field $u_{T\to IT}$ and the residual deformation field $u_{IT\to S}$. We determine $u_{T\to IT}$ automatically with the help of the regression model learnt by statistical models of image appearances and deformation coefficients. To build the statistical models, we first register (using HAMMER, or diffeomorphic demons) a number of training images, to obtain a set of training deformation fields (**Fig. 1(a)**). PCA is then employed on the deformation fields to capture the principal modes of brain deformations (**Fig. 1(b)**). A set of brain appearances (**Fig. 1(c)**), obtained by transforming the training images to the template space, are inverse transformed to the subject

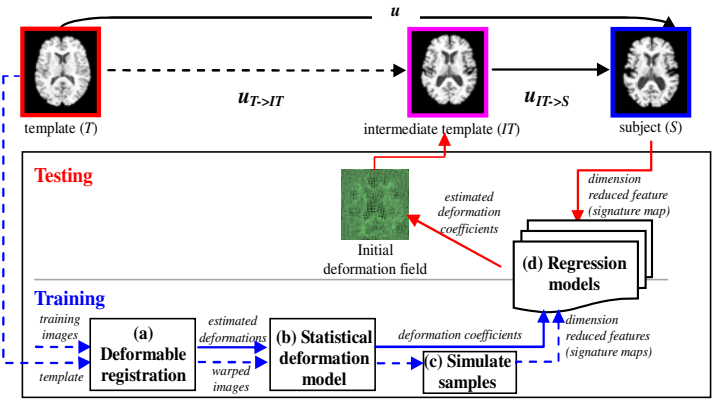


Fig. 1. Schematic illustration of the proposed registration framework

space by the deformation fields generated via perturbing the PCA coefficients (for generating a large number of samples for training). Using these samples, SVR models, bridging the intrinsic appearance-deformation statistics, are trained (**Fig. 1(d)**). Then, given a new subject, an initial deformation field, as well as the corresponding intermediate template, can be instantly predicted by the trained SVR models (**Fig. 1(d)**). Finally, the estimation of residual deformations from the intermediate template to the new subject can be performed at a significantly reduced computation cost due to their similarity. It is worth noting that all existing registration methods can be used for the preparation of training deformation fields (**Fig. 1(a)**), and for refining the registration from the intermediate template to the new subject $u_{IT \to S}$. Details on statistical deformation and appearance models, sample simulation, and construction of SVR models will be given in the following subsections.

2.1 Statistical Deformation-Appearance Model for Sample Simulation

A sufficiently large set of training samples, covering an adequate space of brain shape variation, is important for effective training of the SVR models. Instead of simulating deformation fields based only on some statistical deformation models, as was previously done [6], we design a statistical model which incorporates both deformation and appearance information for simulating training samples.

Statistical deformation model: Given a set of M brain images I_0, their respective deformation fields U_0 can be estimated by a deformable registration algorithm such as HAMMER or diffeomorphic demons. By applying PCA on U_0, we can obtain a statistical deformation model to capture brain shape variations. Specifically, the eigenvectors of deformation covariance matrix represent the principal modes of variation, and their eigenvalues indicate the magnitude of deformation variation along the direction of each corresponding eigenvector. To characterize the principal shape changes, the eigenvectors with the top Q largest eigenvalues are used for approximate representation of the original deformations. The reconstructed counterpart $\hat{u}_{i,0}$ of $u_{i,0} \in U_0$ can be written as:

$$\hat{u}_{i,0} \approx \bar{u} + \sum_{q=1}^{Q} c_{i,0}^{q} \sqrt{\lambda_u^q} \Phi_u^q \,, \tag{1}$$

where $\bar{u}$ denotes the mean deformation field, and λ_u^q and Φ_u^q are the eigenvalues and eigenvectors of the deformation covariance matrix. Each deformation field $u_{i,0}$, is represented in a subspace spanned by $Q$the eigenvectors using a deformation coefficient (column) vector $\vec{c}_{i,0} = \left[c_{i,0}^1 \ldots\ c_{i,0}^q \ldots c_{i,0}^Q\right]^T$, where $c_{i,0}^q = (\Phi_u^q)^T (u - \bar{u}) / \sqrt{\lambda_u^q}$.

Given that we are often faced with the problem of having a limited number of training samples in I_0, we simulate additional brain images based on the statistical model of U_0, to increase the robustness of estimating the brain appearance model. We generate N new deformation coefficient vectors $\vec{c}_{i,j}$ $(i = 1 \ldots M, j = 1 \ldots N)$ for each coefficient vector $\vec{c}_{i,0}$ by perturbing it with $\vec{p}_{i,j}$, i.e., $\vec{c}_{i,j} = \vec{c}_{i,0} + \vec{p}_{i,j}$. The range of each element of the perturbation vector $\vec{p}_{i,j}$ is determined from the mean and standard deviation of the training deformation fields. Thus, a total of $M(N+1)$ simulated deformation fields $\tilde{u}_{i,j}$ can be generated as below:

$$\tilde{u}_{i,j} = \bar{u} + \sum_{q=1}^{Q} \left(c_{i,0}^q + p_{i,j}^q\right) \sqrt{\lambda_u^q} \Phi_u^q. \tag{2}$$

Statistical brain appearance model: To build a more efficient statistical model which is capable of estimating deformation coefficients based on image features, we further *enlarge* the sample dataset by incorporating various brain appearances. Specifically, we first align a set of M training images I_0 onto the template space to get a set of M warped images in the template space. We then generate a series of new brain images $\tilde{I}_{i,k}$ $(i = 1 \ldots M, k = 1 \ldots M(N+1))$ by inverse deforming those M warped images from the template space to the individual spaces by the $M(N+1)$ deformation fields generated by Eq. 2. By including M initial training samples, our statistical model can generate a total of $M^2(N+1)$ samples, thus able to characterize a wide spectrum of deformation and image appearance.

In order to obtain an appearance model feasible for training the regression models, we employ a few strategies to reduce the dimensionality of $\tilde{I}$ before building the appearance model. For each sample image $\tilde{I}_{i,k} \in \tilde{I}$, the background voxels will be first cropped away and then the remained image will be down-sampled, obtaining $\tilde{I}'_{i,k}$. To better represent the shape variation of each sample image $\tilde{I}'_{i,k}$, we further extract brain boundaries along the interfaces between WM, GM, and CSF as the shape descriptors to construct a signature image $SI_{i,k}$. Next, we apply PCA to each signature image $SI_{i,k}$ to represent it by a (low) D-dimensional column vector $\overrightarrow{sm}_{i,k}$, called as *signature map* (SM).

2.2 Construction of SVR Models

After obtaining the deformation coefficient vectors $\vec{c}_{i,k}$ and their corresponding signature maps $\overrightarrow{sm}_{i,k}$, we start to learn their correlations by support vector regression (SVR) models. Specifically, we train Q SVR models, with each responsible for learning the non-linear regression of each row of the deformation coefficient matrix

$C_{Q\times(M+MN)} = [\ldots \vec{c}_{i,k} \ldots]$ with respect to the signature maps $SM_{D\times(M+MN)} = [\ldots \overrightarrow{sm}_{i,k} \ldots]$. SVR is a supervised learning technique for finding non-linear mapping functions which correlate a number of input variables (*features*) to a continuous output variable (*targets*). The features in our case are therefore the signature maps $SM_{D\times(M+MN)}$ and the targets are the rows of the deformation coefficient matrix $C_{Q\times(M+MN)}$. We note here that a common set of signature maps SM is used as the features for all regression models, while the q^{th} row of deformation coefficient matrix C is used independently as the targets for the q^{th} regression model.

For the nonlinear regression, we use a radial basis function (RBF) as the kernel function, and optimize several SVR meta-parameters for building more reliable regression models. We first estimate the kernel size based on the distribution of each signature map $\overrightarrow{sm}_{i,k}$, i.e., the average of the distances from all possible pairs of $\overrightarrow{sm}_{i,k}$. To achieve the global minimum with reliable generalization bound of the regression function, we optimize the parameters ε and γ. Parameter ε is used to control the width of the insensitive zone which penalizes the training data outside this zone. Constant γ (> 0) determines the trade-off between the flatness of the regression function and the tolerance to deviation larger than ε. The value of ε and γ can be calculated as [7]:

$$\varepsilon = \tau \cdot \sigma_s \sqrt{\ln(M + MN)/(M + MN)} \tag{3}$$

$$\gamma = \max(|\bar{c} + 3\sigma_c|, |\bar{c} - 3\sigma_c|), \tag{4}$$

where σ_s is the standard deviation of distances between all pairs of $\overrightarrow{sm}_{i,k}$, τ is an empirical constant, and $M + MN$ is the number of signature maps used as features for regression. $\bar{c}$ and σ_c are the mean and the standard deviation of the deformation coefficients $\vec{c}_{i,k}$, respectively.

2.3 Deformation Prediction for New Subject

After training the regression models, a good initial deformation can be rapidly predicted for any given subject, to bring the template close to the given subject and thus facilitate fast and robust registration. Specifically, for a given subject S, its signiture map $\overrightarrow{sm}_S$ will be first computed by projection onto the Q top eigenvectors (Section 2.2) after affine alignment to the template by FSL FLIRT. The deformation coefficients $\vec{c}_s = (c_s^1\ c_s^2 \ldots c_s^Q)$ can then be predicted one by one through each SVR model. Next, it is straightforward to obtain the initial deformation $u_{T\to IT}$ for subject S by Eq. 1 and also the corresponding intermediate template by warping the template with respect to $u_{T\to IT}$. Now we only need to estimate the residual deformation from the intermediate template to the subject, instead of the original template to the subject as done in many conventional registration algorithms. This step helps save a significant amount of computation cost for overall registration, and also circumvents the error-prone approach of estimating large deformations from the original template to the subject. After estimating $u_{IT\to S}$ (by a conventional registration method, i.e., HAMMER or diffeomorphic demons), $u_{T\to IT}$ and $u_{IT\to S}$ are concatenated to form a final deformation field, from the template to the subject.

3 Experimental Results

We use both HAMMER and diffeomorphic demons for evaluation of our framework and demonstration of its generality. Note that our intention here is not to compare HAMMER and diffeomorphic demons.

3.1 Deformation Prediction

We select randomly $M = 50$ images from a MR brain image database for training. By using both HAMMER and diffeomorphic demons, we first estimate the deformation fields of all these images with respect to a selected template (**Fig. 2a**). By applying PCA, we can build a statistical deformation model with 49 modes from the 49 eigenvectors with non-zero eigenvalues. For denser sampling of the deformation field space, we apply $N = 4$ perturbations, thus giving us a total of $M(N + 1) = 250$ deformation samples, and $M^2(N + 1) = 12500$ brain image samples by inverse deforming those 50 aligned images from the template space to the individual spaces. We then use half (i.e., 6250) of the deformation-image pairs for training the SVR models, and the other half for testing. The averaged prediction error for all deformation coefficients on the training and testing data is 0.1% and 7.8%, respectively. Since the testing data is not used for training, its error is larger than the training error, which is reasonable. This result also indicates that our model is able to predict very good initial deformations for the testing data, which can significantly help reduce the computation cost and registration robustness as reported below.

Given a new subject, we can align it linearly to the template space, and further build its signature map by down-sampling, feature extraction and dimensionality reduction. Then, we can use the SVR models, constructed in the training stage, to predict the initial deformation for the subject. To evaluate the quality of the predicted deformations, we use 50 new images (not used for training) to show their predicted intermediate templates in **Fig. 2.** These results indicate that our method can predict very good

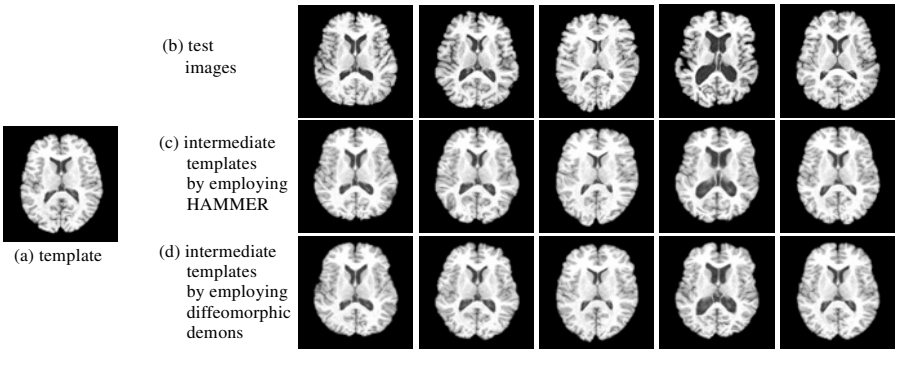
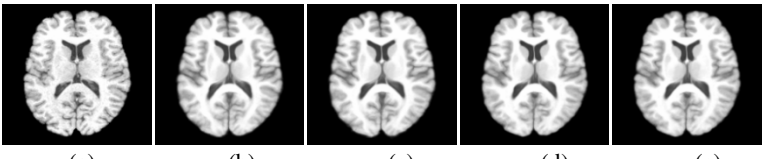


Fig. 2. Demonstration of the intermediate templates (c) and (d) estimated for 5 test images (b), by using HAMMER and diffeomorphic demons, respectively. Compared to the original template (a), the intermediate templates by both methods are very close to the test images, implying that our framework has generality for incorporation of different registration methods.

intermediate templates for both HAMMER and diffeomorphic demons algorithms, conforming its generality for incorporation of different registration methods.

3.2 Registration Performance

Twenty images are separately registered using HAMMER and diffeomorphic demons with and without deformation initialization by our method. The registered images are then averaged to give a visual indication of the registration accuracy, as shown in **Fig. 3**. As can be observed, the average brain images yielded by the algorithms with and without our framework (i.e., (b) and (c) for HAMMER, and (d) and (e) for diffeomorphic demons) are comparable. We note especially that the average brain given by HAMMER using our framework is slightly sharper than that of the original HAMMER, especially in the cortical regions.



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

Fig. 3. Average brain images constructed from 20 brain images aligned by HAMMER and diffeomorphic demons with and without our framework. (a) template; (b) by HAMMER; (c) HAMMER with our framework; (d) diffeomorphic demons; and (e) diffeomorphic demons with our framework.

The registration accuracy is further evaluated using images from NIREP database. Since gray matter regions of brain MRIs in this database were manually segmented into 32 ROIs, we use them to compute the average ROI overlap ratio after registration for accuracy evaluation. However, this database contains only 16 MRI brain images, and we therefore include more images from the OASIS database for training. We select one image from NIREP as the template and register 50 images from OASIS to the selected template, using HAMMER and diffeomorphic demons, respectively, for training our model. Then, we use the remaining 15 images from NIREP for testing the registration performance, by computing the average overlap ratio of the ROIs between the template and each of the 15 aligned images. The average overlap ratio of each ROI is shown in **Fig. 4**. The overall overlap ratio yielded by HAMMER alone is 66%, and with use of our framework its performance is slightly increased to 67%. The overall overlap ratio given by diffeomorphic demons is 64%, and with used of our framework its performance is increased to 66%. These results show that the use of our framework can slightly improve the registration accuracy, with significant speed improvement as mentioned below. We also note that, although the training and testing sets come from different sources, our proposed method is still capable of yielding good performance.

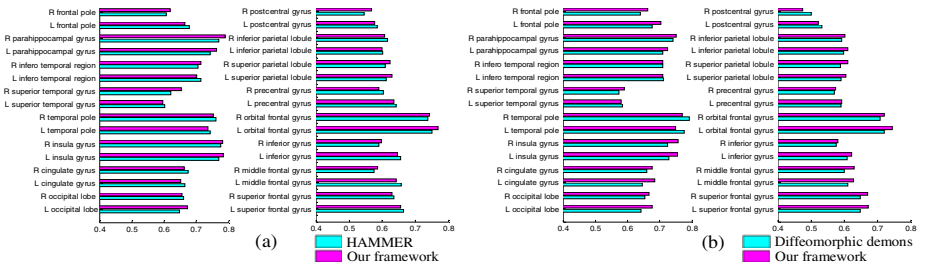


Fig. 4. Average overlap ratios of aligned ROIs. (a) HAMMER with and without our framework, and (b) diffeomorphic demons with and without our framework.

3.3 Speed

Lower computation cost requirement is one of the key advantages of our proposed framework. We use 50 test images of size 256x256x198 to compare the average deformation estimation time of HAMMER and diffeomorphic demons, with and without use of our framework. For similar registration accuracy, HAMMER alone took 90 minutes, and the computation time is reduced to 16 minutes with use of our framework. This indicates about six fold of speed improvement. Diffeomorphic demons alone took 219 seconds, and our framework can help reduce the overall computation time by threefold to 71 seconds.

4 Conclusion

A fast deformable brain registration framework using a novel deformation prediction model has been presented. Specifically, regression models are trained to capture the correlations between image appearances and deformation coefficients. The learnt correlation models are then used to predict rapidly a good initial deformation, as well as the corresponding intermediate template, for any given image. Since the shape difference between the intermediate template and the given image becomes small, the conventional registration methods (i.e., HAMMER and diffeomorphic demons) when incorporated into our framework can perform much faster with comparable accuracy. Future work includes testing more registration methods in the proposed framework such as the B-spline based registration methods.

References

[1] Shen, D., Davatzikos, C.: HAMMER: Hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging 21, 1421–1439 (2002)
[2] Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45, S61–S72 (2009)
[3] Xue, Z., Shen, D., Davatzikos, C.: Statistical Representation of High-Dimensional Deformation Fields with Application to Statistically-Constrained 3D Warping. Medical Image Analysis 10, 740–751 (2006)

[4] Rueckert, D., Frangi, A., Schnabel, J.: Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration. IEEE Transactions on Medical Imaging 22, 1014–1025 (2003)
[5] Loeckx, D., Makes, F., Vandermeulen, D., Suetens, P.: Non-rigid image registration using a statistical spline deformation model. In: Taylor, C.J., Noble, J.A. (eds.) IPMI 2003. LNCS, vol. 2732, pp. 463–474. Springer, Heidelberg (2003)
[6] Tang, S., Fan, Y., Wu, G., Kim, M., Shen, D.: RABBIT: rapid alignment of brains by building intermediate templates. Neuroimage 47, 1277–1287 (2009)
[7] Cherkassky, V., Ma, Y.: Selection of meta-parameters for support vector regression. In: Dorronsoro, J.R. (ed.) ICANN 2002. LNCS, vol. 2415, pp. 687–693. Springer, Heidelberg (2002)

Tracking Clathrin Coated Pits with a Multiple Hypothesis Based Method*

Liang Liang, Hongying Shen, Pietro De Camilli, and James S. Duncan

Yale University, New Haven, CT 06520, USA
liang.liang@yale.edu

Abstract. Cellular processes are crucial for cells to survive and function properly. To study their underlying mechanisms quantitatively with fluorescent live cell microscopy, it is necessary to track a large number of particles involved in these processes. In this paper, we present a method to automatically track particles, called clathrin coated pits (CCPs), which are formed in clathrin mediated endocytosis (CME). The tracking method is developed based on a MAP framework, and it consists of particle detection and trajectory estimation. To detect particles in 2D images and take account of Poisson noise, a Gaussian mixture model is fitted to image data, for which initial parameters are provided by a combination of image filtering and histogram based thresholding methods. A multiple hypothesis based algorithm is developed to estimate the trajectories based on detection data. To use the current knowledge about CCPs, their properties of motion and intensity are considered in our models. The tracking method is evaluated on synthetic data and real data, and experimental results show that it has high accuracy and is in good agreement with manual tracking.

1 Introduction

Clathrin mediated endocytosis (CME) [1] is an essential cellular process that cells use to take up nutrients, to internalize plasma membrane proteins, and to recycle lipid components on the plasma membrane. The study of the process is important in fundamental biological research and virology. It has been found out that the dysfunctions of the process in neurons are correlated to several diseases [1], and CME is one of the major pathways through which viruses enter cells [2]. To understand the process quantitatively, it is necessary to track a large number of particles formed in the process, called clathrin coated pits (CCPs), and obtain their statistics. Since manual tracking is infeasible for large datasets, automatic tracking is important for quantitative studies.

The process can be divided into several stages [1] as illustrated in Fig. 1: clathrin coat assembly, clathrin coat maturation, clathrin coated pits fission into clathrin coated vesicles (CCVs), and finally vesicles uncoating clathrin. Recent advancement in fluorescent live cell microscopy, e.g., spinning disc confocal microscopy (SDCM) and total internal reflection florescent microscopy (TIRFM),

* This work is supported by Keck Foundation.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 315–322, 2010.

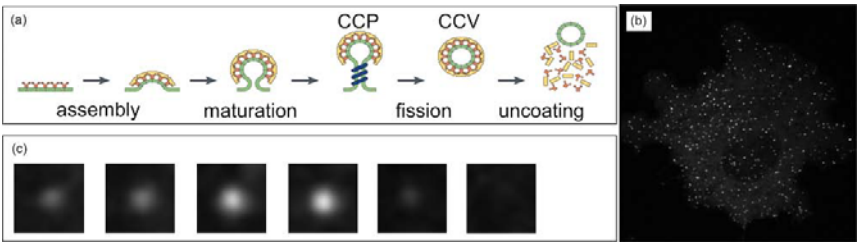


Fig. 1. (a) Different stages of CME. (b) An image taken with SDCM. (c) A sequence showing a CCP in different stages.

enables us to observe CCPs till fission at single particle resolution. As shown in Fig. 1, there are innumerable CCPs on the cell membrane.

Tracking consists of detecting CCPs in each 2D image and estimating the trajectories based on detection data. Although numerous methods for general point tracking have been proposed [3,4], some assumptions on object motion may not be valid for fluorescent particles. Many methods for tracking fluorescent particles have been developed [5,6,7,8,9,10]. In these methods, fluorescent particles are assumed to move with a nearly constant or slow varying velocity, or undergo free diffusion, and their intensities are only related to imaging conditions or their positions. Recently a method [10] is reported for CCP tracking, whose main focus is on data association algorithms. CCPs are not featureless points in images and have their own properties different from other fluorescent particles. CCPs can be created at any time but the rate is limited. Each CCP is connected to cell membrane and can not move freely, and its intensity changes in different stages. To achieve a high tracking accuracy, these properties must be considered.

We present an automatic tracking method based on a MAP framework. A constrained Brownian model is proposed for CCP motion. A linear Gaussian model is used to describe CCP intensity over time. For detection, several methods are used to find reliable positions and intensities of CCPs in each image. A multiple hypothesis based algorithm is developed to find the best trajectories.

2 Method

2.1 The Tracking Framework

Let I_t be the image at time t (frame index), S_t be the joint state of all CCPs at time t, and T be the total number of images. The goal is to find the set of joint states that maximizes a posterior probability:

$$\left\{\widehat{S}_1, ..., \widehat{S}_T\right\} = \arg \max_{S_1 \ to \ S_T} p\left(S_1, ..., S_T | I_1, ..., I_T\right) \tag{1}$$

Assuming CCPs are statistically independent of each other and the process is Markovian, then we obtain

$$p\left(S_1,...,S_T|I_1,...,I_T\right) \propto p\left(S_1\right)\prod_{t=2}^{T} p\left(S_t|S_{t-1}\right)\prod_{t=1}^{T} p\left(I_t|S_t\right) \tag{2}$$

$$p\left(S_t|S_{t-1}\right) = \prod_{i=1}^{N} p\left(X_t^i|X_{t-1}^i\right) \tag{3}$$

X_t^i is the state of the CCP associated with the i^{th} trajectory at time t, and it consists of position, fluorescent intensity and its derivative. The joint state $S_t = \left[X_t^1, X_t^2, ..., X_t^N\right]$, and N is the upper bound on the number of CCP trajectories. $p\left(S_1\right)$ is assumed to be uniform distribution. $p\left(I_t|S_t\right)$ and $p\left(X_t^i|X_{t-1}^i\right)$ will be discussed in section 2.2 and section 2.3 respectively.

Since it is difficult to find the optimal solution of Eq.(1) directly when the CCP number is large, we adopt a conventional strategy, i.e., trajectory estimation after detection. Detection is to find CCP positions and intensities that maximize $p\left(I_t|S_t\right)$ at each time t. Trajectory estimation is to find the correspondences of CCPs in different frames that maximize the product $\prod_{t=2}^{T} p\left(S_t|S_{t-1}\right)$. Detections are performed frame by frame. A multiple hypothesis based algorithm is developed to find the best trajectories given the detection data.

2.2 Detection

In biological experiments, to visualize the dynamics of CCPs, proteins of interest (e.g., clathrin or AP-2 complex in each CCP [1]) are fluorescently labeled, and SDCM or TIRFM is used to obtain time lapse images. The size of each CCP is comparable to the size of the diffraction limited airy disk. As a result, the intensity distribution of each CCP can be described by the point spread function (PSF) which is well approximated by a Gaussian function [11]. During the image capture process in the CCD camera, several types of noises are generated [5]. The major one is shot noise [5] which follows a Poisson distribution and is assumed to be independent at each pixel. Here, we drop the time index of each variable for simplicity. Let F be the fluorescence image without noises, and b be the background intensity level, then we obtain

$$F_{(x,y)} = \sum_{k=1}^{K} f^{(k)} \exp\left(-\frac{\left(x-x^{(k)}\right)^2+\left(y-y^{(k)}\right)^2}{2\sigma^2}\right) + b \tag{4}$$

where $\left(x^{(k)}, y^{(k)}\right)$ and $f^{(k)}$ are position and intensity of the detected particle (CCP) k in the image F.

The detection is to find the set of variables $\left\{b, x^{(k)}, y^{(k)}, f^{(k)}, k = 1, ..., K\right\}$ that maximize the probability $p\left(I|S\right)$ in Eq.(2) at each time t, which is given by

$$p\left(I|S\right) = Poisson\left(I|F\right) = \prod_{(x,y)} \frac{F_{(x,y)}^{I_{(x,y)}} e^{-F_{(x,y)}}}{I_{(x,y)}!} \tag{5}$$

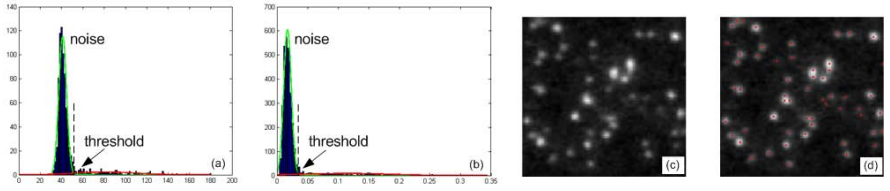


Fig. 2. TIRFM is used and clathrin is fluorescently labeled. (a) Histogram of an original image with fitted distributions. (b) Histogram of the local maxima on the filtered image with fitted distributions. (c) A cropped region. (d) The positions of detected CCPs in the region.

By taking the logarithm, it is equivalent to find the minimum of the function:

$$E_{\left(b,x^{(k)},y^{(k)},f^{(k)},k=1,\ldots,K\right)} = \sum_{(x,y)} \left(F_{(x,y)} - I_{(x,y)} \log\left(F_{(x,y)}\right)\right) \tag{6}$$

The optimal solution can be obtained by using gradient based optimization. Mixture model fitting has been used by some methods [6] to detect fluorescent particles, for which Gaussian image noise is assumed. Here, we use Poisson noise model that fits the image noise well. The background intensity level is estimated by the mean of background intensities. To determine the number of CCPs, we adopt a bottom-up scheme similar to the approach in [6]. To obtain initial values of CCP positions and intensities, we use several methods as shown in Fig. 3. First, local maxima are located by using normalized Laplacian of Gaussian (LoG) filter. Many of them are induced by noises. To separate signals from noises, Gaussian mixture models with two components are fitted to the histograms of the original image and the local maxima (in the filtered image) by invoking EM algorithm. After thresholding, the surviving local maxima give the initial values.

2.3 Motion and Intensity Modeling

X_t is the state of a CCP at time t and $X_t = [x_t, y_t, f_t, g_t]'$. (x_t, y_t) is the position in the x-y plane, f_t is the fluorescence intensity, and g_t is the derivative of f_t with respect to t. Here, we drop the trajectory-index of each variable for simplicity. $p\left(X_t|X_{t-1}\right)$ in Eq.(3) is the state evolution model and can be factorized as

$$p\left(X_t|X_{t-1}\right) = p\left(x_t, y_t|x_{t-1}, y_{t-1}\right) p\left(f_t, g_t|f_{t-1}, g_{t-1}\right) \tag{7}$$

$p\left(x_t, y_t|x_{t-1}, y_{t-1}\right)$ is the motion model, and $p\left(f_t, g_t|f_{t-1}, g_{t-1}\right)$ is the model of fluorescence intensity over time. The factorization is based on the plausible assumption that intensities are independent with x-y positions for each CCP.

Motion Modeling: CCP motion is mainly caused by two factors. First, tiny molecules in cytosol randomly bombard CCPs, which causes CCPs to move. Second, the forces induced by clathrin coat assembly will cause vibrations of CCPs. Since each CCP is connected to the plasma membrane, it can only move within

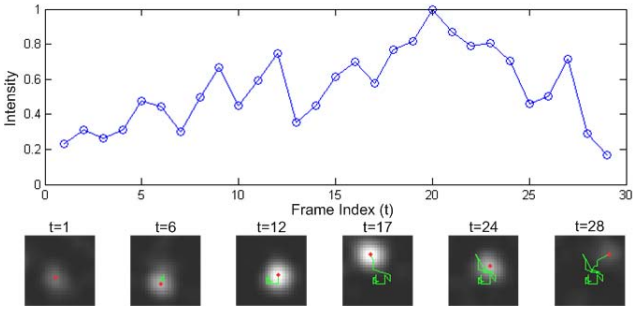


Fig. 3. TIRFM is used and AP-2 complex is fluorescently labeled. (a) The intensity-time curve of the CCP. (b) Each red dot indicates the position at each time t, and each green curve shows the trajectory up to each time t.

a restricted region as shown in Fig. 3. Therefore, we propose the constrained Brownian motion model to describe CCP motion, given by

$$p\left(x_t, y_t | x_{t-1}, y_{t-1}\right) \propto \exp\left(-\frac{\left(x_t - x_{t-1}\right)^2 + \left(y_t - y_{t-1}\right)^2}{2\sigma_m^2}\right) \exp\left(-\frac{\left(x_t - x_c\right)^2 + \left(y_t - y_c\right)^2}{2\sigma_c^2}\right) \quad (8)$$

where $x_c = \frac{1}{t-t_1+1}\sum_{\tau=t_1}^{t} x_\tau$ and $y_c = \frac{1}{t-t_1+1}\sum_{\tau=t_1}^{t} y_\tau$. t_1 is the starting time of the trajectory. σ_m and σ_c are estimated from training data.

Intensity Modeling: CCP intensity changes over time. Fig. 3 shows a typical intensity-time curve. The gradual increase of fluorescence intensity corresponds to CCP creation and growth. The decrease indicates that the connection between the CCP and cell membrane will be cut off, and then it will disappear quickly in the image. Therefore, intensity over time is directly modeled by using a linear Gaussian model, given by

$$p\left(f_t, g_t | f_{t-1}, g_{t-1}\right) \propto \exp\left(-\frac{1}{2}\left[f_t - f_{t-1}, g_t - g_{t-1}\right] Q^{-1} \left[f_t - f_{t-1}, g_t - g_{t-1}\right]'\right) \quad (9)$$

where Q is learned from training data.

2.4 Trajectory Estimation

There are many general methods [12] for solving the correspondence problem given the detection data. We use the multiple hypothesis approach [13] due to its flexibility, and tailor it to our application.

Suppose the CCP trajectory i starts from time (frame) t_1 and ends at time t_2. The lifetime of the CCP is $t_2 - t_1 + 1$. The cost of the trajectory is defined as $C^i = -\sum_{t=2}^{T} \log p\left(X_t^i | X_{t-1}^i\right)$.

The cost of the correspondence between the CCP trajectory i in frame $t-1$ and the detected CCP k in frame t is defined as $C(i,k) = -\log p\left(X_t^{(k)}|X_{t-1}^i\right)$. Here, $X_t^{(k)}$ is the k^{th} "candidate" for state X_t^i, and $X_t^{(k)} = \left[x_t^{(k)}, y_t^{(k)}, f_t^{(k)}, g_t^{(k)}\right]'$. Measurements of the states are provided by the detection module discussed in section 2.2. Since the key biological parameter is CCP lifetime, measurement noises of positions and intensities can be ignored.

To deal with CCP appearing, set $X_{t_1-1}^i = \left[x_{t_1}^i, y_{t_1}^i, b_{t_1} + \gamma\sigma_{b(t_1)}, 0\right]'$, which means the CCP's intensity is low when it is created. To deal with CCP disappearing, set $X_{t_2+1}^i = \left[x_{t_2}^i + \Delta x, y_{t_2}^i + \Delta y, b_{t_2} + \gamma\sigma_{b(t_2)}, 0\right]'$, which means it will leave the current position with a low intensity. b_t is the background intensity level, and $\sigma_{b(t)}$ is the standard deviation of background intensities. Δx and Δy are set to $\eta\sigma_m$. Multiplication factors (γ and η) are learned from training data. For $t \in [1, t_1 - 1] \cup [t_2 + 2, T]$, set $p\left(X_t^i|X_{t-1}^i\right)$ =constant(> 0), which means the states are irrelevant when the CCP has not been created or has disappeared.

By using Eq.(3), the total cost is defined as

$$C_{total} = -\log\left(\prod_{t=2}^{T} p\left(S_t|S_{t-1}\right)\right) = -\sum_{i=1}^{N}\sum_{t=2}^{T}\log p\left(X_t^i|X_{t-1}^i\right) = \sum_{i=1}^{N} C^i \quad (10)$$

With these cost functions, multiple hypotheses can be generated and pruned to find the correspondences associated with the minimal total cost. We develop an algorithm based on MHT [13]. The generation of hypotheses is achieved by random sampling according to the soft-assign matrices [14].

3 Experimental Results

3.1 Evaluation on Synthetic Data

The proposed method is evaluated on synthetic 2D image datasets. Each dataset is generated from a noise-free image sequence of moving particles (200 images of 120x120 pixels) by adding different level of noises, and the number of particles (CCPs) is 386. To simulate CCP motion, we fit Gaussian distributions to the histograms of displacements and deviations from the mean positions, and therefore trajectories can be generated by sampling from these fitted distributions. Each simulated CCP has a linear intensity-time curve. We choose exponential distribution as the lifetime distribution based on the current knowledge about CCPs. If a CCP's intensity is near the background level, it may not be discriminated from noise. Therefore, the SNR of a dataset is defined as

$$SNR = \frac{\text{Average CCP Intensity } - \text{Average Background Intensity}}{\text{Standard Deviation of Background Intensities}}$$

Since the key parameter is CCP lifetime, tracking accuracy is defined as

$$Accuracy = \frac{\sum_k (\text{Number of Correct Associations for the Trajectory } k)}{\sum_k (\text{True Lifetime of the CCP with the Trajectory } k)}$$

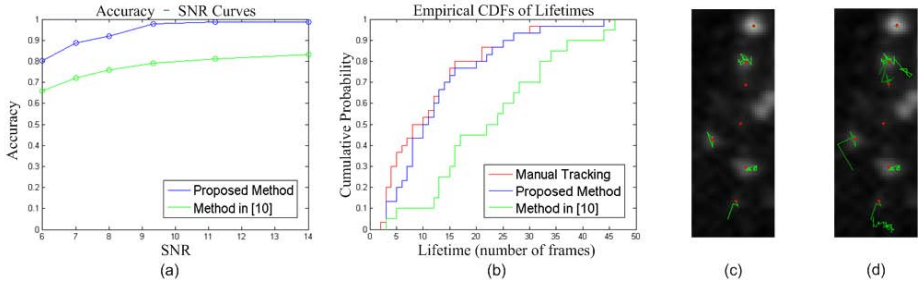


Fig. 4. (a) Accuracy-SNR curves. (b) Empirical cumulative distribution functions (CDFs) of lifetimes. (c) Samples of trajectories obtained by using the proposed method. (d) Samples of trajectories obtained by using the trajectory estimation module of the method in [10]. Red dots indicate the positions of the CCPs at t=33 and green curves show their trajectories up to the time t.

The definition is similar to the metric in [15]. The best match between ground truth trajectories and estimated trajectories are found by using the matching algorithm in [15]. Supposing that the estimated trajectory k is matched to the ground truth trajectory n, if a detected CCP in any frame is associated with both trajectories, then the estimated trajectory k has a correct association.

We also test the method reported in [10]. It uses pure Brownian motion model which allows CCPs move freely, and it does not consider intensity variations of individual CCPs. For detection, its model fitting module is selected because CCPs are homogeneous in the simulation.

The *Accuracy* - SNR curves are shown in Fig. 4(a). The proposed method is consistently more accurate, most likely in part due to the better models.

3.2 Evaluation on Real Data

COS7 cells were electroporated with clathrin light chain GFP construct using the Amaxa Nucleofector method, and were plated at subconfluent densities into 35mm glass bottom dishes, and allowed to grow for 12 to 48 hours. Then, TIRFM images were acquired using an inverted microscope equipped with a high numerical aperture (NA=1.49, 60X) lens (Olympus) and a back illuminated Andor iXon887 EMCCD camera, controlled by Andor iQ software (Andor Technology).

Since ground truth is not available, 30 CCPs are manually tracked by a human expert biologist, to serve as reference data. To show the effectiveness of the proposed CCP models, we test the trajectory estimation modules of the proposed method and the method in [10] based on the same detection data provided by the method in section 2.2. The results are shown in Fig. 4(b)–(d). Kolmogorov-Smirnov test (KS-test) is used to measure the difference between lifetime distributions. The proposed method has good agreement with manual tracking ($p>0.5$). The alternative method produces longer trajectories ($p<0.01$), which is most likely to happen when new CCPs appear in the vicinities of disappeared CCPs. Thus, better models are helpful to prevent erroneous links.

4 Conclusion

We have proposed a method to automatically track Clathrin Coated Pits (CCPs) in clathrin mediated endocytosis (CME). Starting from a MAP framework, we have developed algorithms for CCP detection and trajectory estimation. Some properties of CCPs are considered in our models, which is different from related works. We also consider the Poisson image noise in the mixture model fitting procedure. The proposed method has been demonstrated on synthetic data and real data. It will be used by the biologists to investigate mechanisms of CME.

References

1. Slepnev, V.I., Camilli, P.D.: Accessory factors in clathrin-dependent synaptic vesicle endocytosis. Nature Reviews Neuroscience 1, 161–172 (2000)
2. Brandenburg, B., Zhuang, X.: Virus trafficking - learning from single-virus tracking. Nature Reviews Microbiology 5, 197–208 (2007)
3. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Journal of Computing Surveys 38(4) (2006)
4. Veenman, C.J., Reinders, M.J.T., Backer, E.: Resolving motion correspondence for densely moving points. IEEE Trans. on Pattern Analysis and Machine Intelligence 23(1), 54–72 (2001)
5. Carter, B.C., Shubeita, G.T., Gross, S.P.: Tracking single-particles: a user-friendly quantitative evaluation. Physical Biology 2, 60–72 (2005)
6. Thomann, D., Rines, D.R., Sorger, P.K., Danuser, G.: Automatic fluorescent tag detection in 3D with super-resolution: application to the analysis of chromosome movement. J. of Microscopy 208(1), 49–64 (2002)
7. Sbalzarini, I.F., Koumoutsakos, P.: Feature point tracking and trajectory analysis for video imaging in cell biology. J. of Structural Biology 151, 182–195 (2005)
8. Yang, G., Matov, A., Danuser, G.: Reliable tracking of large-scale dense particle motion for fluorescent live cell imaging. In: Proc. of IEEE Int. Conf. Computer Vision and Pattern Recognition (2005)
9. Smal, I., Niessen, W.J., Meijering, E.: Advanced particle filtering for multiple object tracking in dynamic fluorescence microscopy images. In: IEEE Int. Symposium on Biomedical Imaging: From Nano to Macro, 1048–1051 (2007)
10. Jaqaman, K., Loerke, D., Mettlen, M., Kuwata, H., Grinstein, S., Schmid, S.L.L., Danuser, G.: Robust single-particle tracking in live-cell time-lapse sequences. Nature methods 5, 695–702 (2008)
11. Zhang, B., Zerubia, J., Olivo-Marin, J.-C.: Gaussian approximations of fluorescence microscope point-spread function models. Applied Optics 46(10), 1819–1829 (2007)
12. Poore, A.B., Gadaleta, S.: Some assignment problems arising from multiple target tracking. Mathematical and Computer Modelling 43, 1074–1091 (2006)
13. Reid, D.B.: An algorithm for tracking multiple targets. IEEE Trans. on Automatic Control 24, 843–854 (1979)
14. Rangarajan, A., Chui, H., Bookstein, F.L.: The softassign procrustes matching algorithm. In: Duncan, J.S., Gindi, G. (eds.) IPMI 1997. LNCS, vol. 1230, pp. 29–42. Springer, Heidelberg (1997)
15. Kasturi, R., et al.: Framework for performance evaluation of face, text, and vehicle detection and tracking in video: data, metrics, and protocol. IEEE Trans. on Pattern Analysis and Machine Intelligence 31(2), 319–336 (2009)

Shape-Based Diffeomorphic Registration on Hippocampal Surfaces Using Beltrami Holomorphic Flow

Lok Ming Lui[1,2], Tsz Wai Wong[2], Paul Thompson[3], Tony Chan[4], Xianfeng Gu[5], and Shing-Tung Yau[1]

[1] Department of Mathematics, Harvard University, Cambridge, MA, USA
[2] Department of Mathematics, UCLA, Los Angeles, CA, USA
[3] Laboratory of Neuro Imaging, UCLA School of Medicine, Los Angeles, CA, USA
[4] Hong Kong University of Science and Technology, Hong Kong
[5] Department of Computer Science, SUNY Stony Brook, Stony Brook, NY, USA

Abstract. We develop a new algorithm to automatically register hippocampal(HP) surfaces with complete geometric matching, avoiding the need to manually label landmark features. A good registration depends on a reasonable choice of shape energy that measures the dissimilarity between surfaces. In our work, we first propose a complete shape index using the Beltrami coefficient and curvatures, which measures subtle local differences. The proposed shape energy is zero if and only if two shapes are identical up to a rigid motion. We then seek the best surface registration by minimizing the shape energy. We propose a simple representation of surface diffeomorphisms using Beltrami coefficients, which simplifies the optimization process. We then iteratively minimize the shape energy using the proposed Beltrami Holomorphic flow (BHF) method. Experimental results on 212 HP of normal and diseased (Alzheimer's disease) subjects show our proposed algorithm is effective in registering HP surfaces with complete geometric matching. The proposed shape energy can also capture local shape differences between HP for disease analysis.

1 Introduction

The hippocampus(HP) is an important subcortical structure of the human brain that plays a key role in long-term memory and spatial navigation. Surface-based shape analysis is commonly used to study local changes of HP surfaces due to pathologies such as Alzheimer disease (AD), schizophrenia and epilepsy[11]. When comparing data on two anatomical surfaces, a 1-1 correspondence must be computed to register one surface nonlinearly onto the other. On HP surfaces, there are no well-defined anatomical landmark features that can be used as a constraint to establish good correspondences. High-field structural or functional imaging, where discrete cellular fields are evident [15], is still not routinely used. Finding meaningful registrations between HP surfaces becomes challenging. Inaccuracies in shape analysis are often introduced due to incorrect registrations. In fact, shape analysis and surface registration are closely related. The results of

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 323–330, 2010.

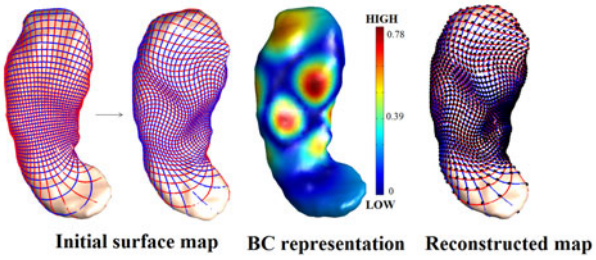


Fig. 1. Representation of surface registration using Beltrami Coefficients

shape analysis can be highly affected by the registration, but a good registration depends largely on the appropriate choice of shape measure that captures dissimilarities between surfaces. Therefore, it is of utmost importance to combine the two processes and define a suitable shape measure to drive the registration.

Here we developed an algorithm to automatically register HP surfaces with complete geometric matching, avoiding the need to manually label landmark features. We first propose a complete shape index using the Beltrami coefficient (BC) and curvatures, which measures subtle local differences. The shape energy is identically zero if and only if two shapes are equal up to a rigid motion. We then minimize the shape energy to obtain the best surface registration with complete geometric matching. We propose a simple representation of surface diffeomorphisms using BCs, which simplifies the optimization. We then optimize the shape energy using the Beltrami Holomorphic flow (BHF) method. The optimal shape energy obtained may also be used to measure local shape differences across subjects or time.

2 Related Work

Surface registration has been studied extensively. Conformal or quasi-conformal surface registration is commonly used [4,5,14], and gives a parameterization minimizing angular distortions. However, it cannot guarantee the matching of geometric information such as curvature across subjects. Landmark-based diffeomorphisms are often used to compute, or adjust, cortical surface parameterizations [3,6,12]. These methods provide good registrations when corresponding landmark points on the surfaces can be labeled in advance. It is, however, difficult for HP surfaces on which there are no well-defined anatomical landmarks. Some authors have proposed driving features into correspondence based on shape information. Lyttelton et al. [8] computed surface parameterizations that match surface curvature. Fischl et al. [1] improved the alignment of cortical folding patterns by minimizing the mean squared difference between the average convexity across a set of subjects and that of the individual. Wang et al. [13] computed surface registrations that maximize the mutual information between mean curvature and conformal factor maps across subjects. Lord et al. [7] matched surfaces by minimizing the deviation from isometry. The shape indices that drive the

registration process in these approaches are not complete shape measurements and do not capture shape differences completely. There are cases when two different surfaces might have the same shape value. This could lead to inaccurate registration results.

3 Theoretical Background and Definitions

Given two Riemann surfaces M and N, a map $f : M \rightarrow N$ is *conformal* if it preserves the surface metric up to a multiplicative factor. One generalization of conformal maps is the *quasi-conformal* maps, which are orientation-preserving homeomorphisms between Riemann surfaces with bounded conformality distortion, in the sense that their first order approximations takes small circles to small ellipses of bounded eccentricity [2]. Thus, a conformal homeomorphism that maps a small circle to a small circle may also be regarded as quasi-conformal. Mathematically, $f\colon \mathbb{C} \rightarrow \mathbb{C}$ is quasi-conformal if it satisfies the Beltrami equation: $\frac{\partial f}{\partial \bar{z}} = \mu(z)\frac{\partial f}{\partial z}$, for some complex valued function μ satisfying $||\mu||_\infty < 1$. μ is called the *Beltrami coefficient* (BC), which is a measure of non-conformality. In particular, the map f is conformal around a small neighborhood of p when $\mu(p) = 0$. From $\mu(p)$, we can determine the angles of the directions of maximal magnification and shrinking and the amount of them as well. Specifically, the angle of maximal magnification is $\arg(\mu(p))/2$ with magnifying factor $1 + |\mu(p)|$; The angle of maximal shrinking is the orthogonal angle $(\arg(\mu(p)) - \pi)/2$ with shrinking factor $1 - |\mu(p)|$. The distortion or dilation is given by: $K = (1 + |\mu(p)|)/(1 - |\mu(p)|)$.

4 Proposed Model

4.1 A Complete Shape Index

A good registration depends greatly on the appropriate choice of a shape measure to capture dissimilarities between surfaces. We propose a complete shape index E_{shape} using the Beltrami coefficient and curvatures, which measures subtle local changes completely. Given two HP surfaces S_1 and S_2. Let $f : S_1 \rightarrow S_2$ be a registration between S_1 and S_2. The complete shape index E_{shape} is defined as follow: $E_{shape}(f) = \alpha|\mu|^2 + \beta(H_1 - H_2(f))^2 + \gamma(K_1 - K_2(f))^2$ where μ is the Beltrami coefficient of f; H_1, H_2 are the mean curvatures on S_1 and S_2 respectively; and K_1, K_2 are the Gaussian curvatures. The first term measures the conformality distortion of the surface registration. The second and third terms measure the curvature mismatch. It turns out E_{shape} is a complete shape index that measures subtle shape differences between two surfaces. It can be proven that $E_{shape}(f) = 0$ if and only if S_1 and S_2 are equal up to a rigid motion. For HP shape analysis, it is good because clinically we are more interested in shape changes than their orientation. Also, by adjusting the parameters (i.e., α, β and γ), E_{shape} can be made equivalent to other existing shape indices. For example, when $\beta = 0$, E_{shape} is equivalent to the isometric shape index; when

$\alpha = 0$, E_{shape} is equivalent to the curvature index; when $\beta = \gamma = 0$, E_{shape} measures the conformality distortion. In our work, we set $\alpha = 1$ and $\beta = \gamma = 2$ to measure complete shape changes.

We can now minimize E_{shape} to obtain the optimized surface map $\widetilde{f}$ that best matches the geometry. One advantage of using E_{shape} is that it can be defined in the space of BCs. The space of BCs is a simple functional space, which makes the optimization much easier.

4.2 Surface Map Representation Using Beltrami Coefficients

Surface registration is commonly parameterized using 3D coordinate functions in $\mathbb{R}^3$. This representation is difficult to manipulate. For example, the 3D coordinate functions have to satisfy certain constraints on the Jacobian J (namely, $J > 0$), to preserve the 1-1 correspondence of the surface maps. Enforcing this constraint adds extra difficulty in optimizing surface maps. The diffeomorphic property is often lost during the optimization. We propose a simple representation of surface diffeomorphisms using Beltrami coefficients (BCs). Fixing any 3 points on a pair of surfaces, there is a 1-1 correspondence between the set of surface diffeomorphisms between them and the set of BCs on the source domain.

Suppose S_1 and S_2 are both either genus 0 closed surfaces or simply connected open surfaces. S_1 and S_2 can be conformally parameterized with a global patch D[4,14]. Let $f\colon S_1 \to S_2$, and given 3 point correspondences. In this work, we chose the 3 corresponding points based on the initial conformal registration. But we can easily generalize our method by incorporating a Mobius transformation that will help us to automatically detect optimal 3-point correspondences. Denote the parameterizations by $\phi_1\colon S_1 \to D$ and $\phi_2\colon S_2 \to D$. Now, we can compute the Beltrami coefficient μ_f associated uniquely to f to represent f (See Figure 1). The Beltrami coefficient μ_f can be computed by considering the composition map $\widetilde{f} = \phi_2 \circ f \circ \phi_1^{-1}\colon D \to D$. Mathematically, μ_f is given by the following formula: $\mu_f = \frac{\partial \widetilde{f}}{\partial \overline{z}} / \frac{\partial \widetilde{f}}{\partial z} = \frac{1}{2}(\frac{\partial \widetilde{f}}{\partial x} + \sqrt{-1}\frac{\partial \widetilde{f}}{\partial y}) / \frac{1}{2}(\frac{\partial \widetilde{f}}{\partial x} - \sqrt{-1}\frac{\partial \widetilde{f}}{\partial y})$.

The space of BCs is a simple functional space. There are no restrictions on μ that it has to be 1-1, surjective or satisfy some constraints on the Jacobian. Using the Beltrami representation makes the optimization process of surface maps much easier.

4.3 Optimized Surface Registration Matching the Geometry

E_{shape} gives us a complete shape index which measures local dissimilarities between two surfaces. Specifically, $E_{shape}(f) = 0$ if and only if S_1 and S_2 are equal up to a rigid motion. Therefore, the surface map f minimizing $E_{shape}(f)$ is the best registration that best matches the geometric information. Given two HP surfaces S_1 and S_2. We propose to find $f : S_1 \to S_2$ that minimizes $E = \int E_{Shape}(f)$. To simplify the computation, we can conformally parameterize S_1 and S_2 onto the parameter domain D. So, all computations are carried out on the simple domain D. By representing surface maps with Beltrami coefficients μ,

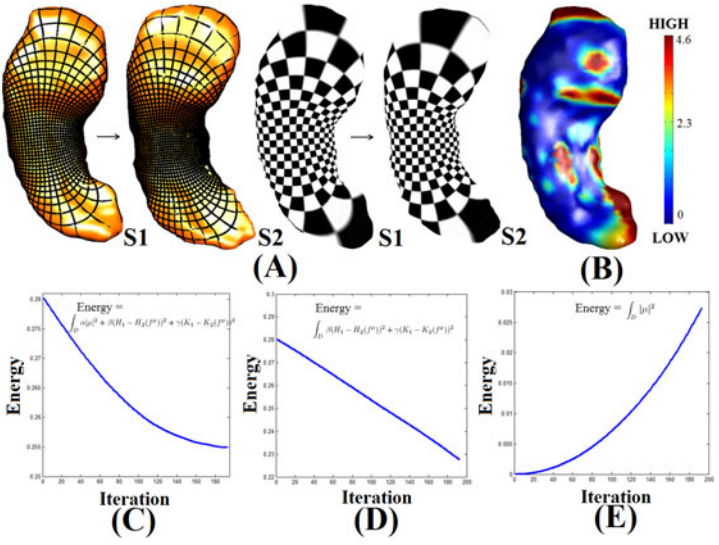


Fig. 2. Shape registration with geometric matching using Beltrami Holomorphic flow

we can define the energy on the space of BCs - a much simpler functional space for the optimization process. Mathematically, the compound energy E can be written with respect to μ as: $E(\mu) = \int_D \alpha|\mu|^2 + \beta(H_1 - H_2(f^\mu))^2 + \gamma(K_1 - K_2(f^\mu))^2$. The variation of f^μ under the variation of μ can be expressed explicitly. Suppose $\widetilde{\mu}(z) = \mu(z) + t\nu(z) + \mathcal{O}(t^2)$. Then, $f^{\widetilde{\mu}(z)}(w) = f^\mu(w) + tV(f^\mu, \nu)(w) + \mathcal{O}(t^2)$, where $V(f^\mu, \nu)(w) = -\frac{f^\mu(w)(f^\mu(w)-1)}{\pi} \int_D \frac{\nu(z)(f_z^\mu(z))^2 dxdy}{f^\mu(z)(f^\mu(z)-1)(f^\mu(z)-f^\mu(w))}$ Using the variational formula, we can derive the Euler-Lagrange equation of $E(\mu)$ easily. Specifically, we can minimize $E(\mu)$ by the following iterative scheme:
$\mu^{n+1} - \mu^n = -2(\alpha\mu^n - \int_z[(\beta\widetilde{H}^n + \gamma\widetilde{K}^n) \cdot G^n, \mathbf{det}(\beta\widetilde{H}^n + \gamma\widetilde{K}^n, G^n)]\ )dt$,
where $\int_w \bullet := \int_D \bullet\, dw$ and $\int_z \bullet := \int_D \bullet\, dz$ is defined as the integral over the variable w and z respectively; $\widetilde{H} := (H_1 - H_2(f^\mu))\nabla H_2(f^\mu)$; $\widetilde{K} := (K_1 - K_2(f^\mu))\nabla K_2(f^\mu)$; $\mathbf{det}(a, b)$ is the determinant of the 2 by 2 matrix or equivalently, the norm of the cross product of a and b.

We call this iterative algorithm the *Beltrami Holomorphic flow* (BHF). Note that starting with a conformal map with $\mu = 0$, the first term of the energy ensures μ to satisfy $||\mu||_\infty < 1$. Hence, during the BHF process, the maps are guaranteed to be diffeomorphic and are holomorphic in t.

5 Experimental Results

We tested our algorithm on 212 HP surfaces automatically extracted from 3D brain MRI scans with a validated algorithm [9]. Scans were acquired from normal and diseased (AD) elderly subjects at 1.5 T (on a GE Signa scanner). Experiments have been carried out on a laptop with a 2.4 GHz DUO CPU. The

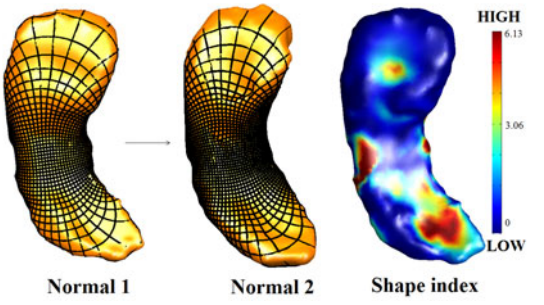


Fig. 3. BHF registration between two normal subjects. The shape index E_{shape} is plotted on the right, which captures local shape differences.

algorithm takes about 4-5 minutes to compute a registration between meshes with 40K vertices.

Figure 1 shows the Beltrami representations of bijective surface maps. The left column shows a bijective surface map between the HP surfaces. The middle column shows the Beltrami (BC) representations of the maps. The right column shows the reconstruction of surface maps from their BCs. The reconstructed maps closely resemble the original maps, meaning that BCs can effectively represent bijective surface maps. Figure 2(A) shows two different HP surfaces. They are registered using our proposed BHF algorithm with geometric matching. The registration is visualized using a grid map and texture map, which shows a smooth 1-1 correspondence. The optimal shape index E_{shape} is plotted as colormap in (B). E_{shape} effectively captures the local shape difference between the surfaces. (C) shows the shape energy in each iteration. With the BHF algorithm, the shape energy decreases as the number of iterations increases. (D) shows the curvature mismatch energy ($E = \int \beta(H_1 - H_2(f))^2 + \gamma(K_1 - K_2(f))^2$). It decreases as the number of iterations increases, meaning that the geometric matching improves. (E) shows the Beltrami coefficient of the map in each iteration, which shows the conformality distortion of the map. Some conformality is intentionally lost to allow better geometric matching.

Figure 3 shows the BHF registration between two normal HPs. The complete shape index E_{shape} is plotted as colormap on the right. Again, E_{shape} can accurately capture local shape differences between the normal HP surfaces.

Figure 4 shows the BHF hippocampal registrations between normal elderly subjects and subjects with Alzheimer's disease. The BHF registrations give smooth 1-1 correspondences between the HP surfaces. We can use the complete shape index E_{shape} to detect local shape differences between healthy and unhealthy subjects.

We also study the temporal shape changes of normal and AD HP surfaces, as shown in Figure 5. For each subject, we compute the deformation pattern of its HP surfaces measured at time = 0 and time = 12 Months (see [10] for longitudinal scanning details). The left two panels show the temporal deformation patterns for two normal subjects. The middle two panels show the temporal

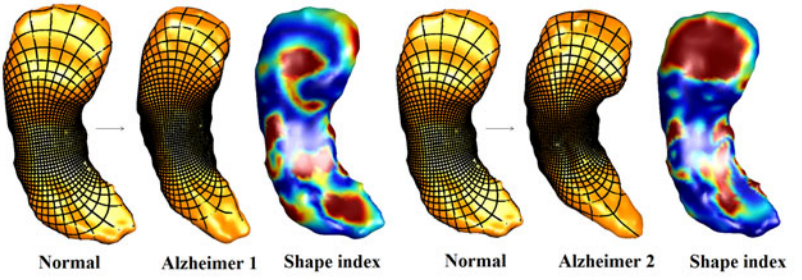


Fig. 4. BHF registration between normal subjects and subjects with Alzheimer's disease. Their local shape differences are captured by E_{shape}.

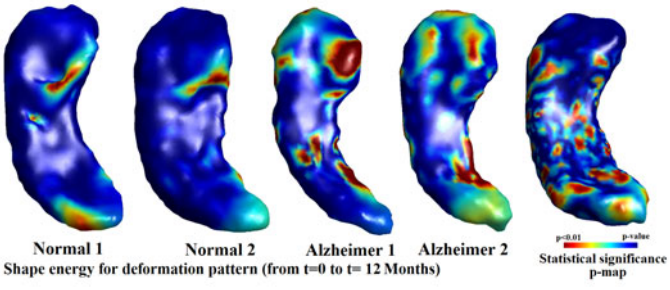


Fig. 5. Temporal hippocampal shape changes of normal and subjects with Alzheimer's disease

deformation patterns for two AD subjects. The last column shows the statistical significance p-map measuring the difference in the deformation pattern between the normal (n=47) and AD (n=53) groups, plotted on a control HP. The deep red color highlights regions of significant statistical difference. This method can be potentially used to study factors that influence brain changes in AD.

6 Conclusion and Future Work

We developed an algorithm to automatically register HP surfaces with complete geometric matching, avoiding the need for manually-labeled landmark features. We did this by defining a complete shape index to drive the registration. Experimental results on 212 HP surfaces from normal and diseased(AD) subjects show our proposed algorithm is effective in registering HP surfaces over time and across subjects, with complete geometric matching. The proposed shape energy can also capture local shape differences between HPs for disease analysis. In future, we will use the BHF algorithm to systematically study the local shape differences and factors that affect deformation patterns between normal and AD subjects.

Acknowledgement. PT is supported by NIH: EB008432, EB008281, EB007813, HD050735 & AG036535. TF is supported by NSF GEO-0610079, NIH

U54 RR021813 and ONR N00014-09-1-0105. XG is supported by NIH 1R01EB0075300A1, NSF IIS 0916286, CCF0916235, CCF0830550, III0713145, and ONR N000140910228.

References

1. Fischl, B., et al.: High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping 8, 272–284 (1999)
2. Gardiner, F., et al.: Quasiconformal Teichmuller Theory. Amer. Math. Soc., Providence (2000)
3. Glaunés, J., et al.: Landmark matching via large deformation diffeomorphisms on the sphere. J. Maths. Imaging and Vision 20, 179–200 (2004)
4. Gu, X., et al.: Genus zero surface conformal mapping and its application to brain surface mapping. IEEE Transactions on Medical Imaging 23(8), 949–958 (2004)
5. Hurdal, M.K., et al.: Discrete conformal methods for cortical brain flattening. Neuroimage 45, 86–98 (2009)
6. Leow, A., et al.: Brain structural mapping using a novel hybrid implicit/explicit framework based on the level-set method. NeuroImage 24(3), 910–927 (2005)
7. Lord, N.A., et al.: Simultaneous registration and parcellation of bilateral hippocampal surface pairs for local asymmetry quantification. IEEE Transactions on medical imaging 26(4), 471–478 (2007)
8. Lyttelton, O., et al.: An unbiased iterative group registration template for cortical surface analysis. NeuroImage 34, 1535–1544 (2007)
9. Morra, J., et al.: Validation of a fully automated 3d hippocampal segmentation method using subjects with alzheimer's disease, mild cognitive impairment, and elderly controls. Neuroimage 43(1), 59–68 (2008)
10. Morra, J., et al.: Automated mapping of hippocampal atrophy in 1-year repeat mri data in 490 subjects with alzheimer's disease, mild cognitive impairment, and elderly controls. Neuroimage, 45(1), S3–S15 (2009)
11. Thompson, P., et al.: Mapping hippocampal and ventricular change in alzheimer's disease. NeuroImage 22(4), 1754–1766 (2004)
12. Thompson, P., et al.: A surface-based technique for warping 3-dimensional images of the brain. IEEE Transactions on Medical Imaging 15(4), 1–16 (1996)
13. Wang, Y., et al.: Automated surface matching using mutual information applied to Riemann surface structures. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 666–674. Springer, Heidelberg (2005)
14. Wang, Y., et al.: Brain surface conformal parameterization using Riemann surface structure. IEEE Transactions on Medical Imaging 26(6), 853–865 (2007)
15. Zeineh, M., et al.: Dynamics of the hippocampus during encoding and retrieval of face-name pairs. NeuroImage 299(5606), 577–580 (2003)

Detecting Brain Activation in fMRI Using Group Random Walker

Bernard Ng[1], Ghassan Hamarneh[2], and Rafeef Abugharbieh[1]

[1] Biomedical Signal and Image Computing Lab, The University of British Columbia
[2] Medical Image Analysis Lab, Simon Fraser University
{Bernardn,rafeef}@ece.ubc.ca, hamarneh@cs.sfu.ca

Abstract. Due to the complex noise structure of functional magnetic resonance imaging (fMRI) data, methods that rely on information within a single subject often results in unsatisfactory functional segmentation. We thus propose a new graph-theoretic method, "Group Random Walker" (GRW), that integrates group information in detecting single-subject activation. Specifically, we extend each subject's neighborhood system in such a way that enables the states of both intra- and inter-subject neighbors to be regularized without having to establish a one-to-one voxel correspondence as required in standard fMRI group analysis. Also, the GRW formulation provides an exact, unique closed-form solution for jointly estimating the probabilistic activation maps of all subjects with global optimality guaranteed. Validation is performed on synthetic and real data to demonstrate GRW's superior detection power over standard analysis methods.

Keywords: fMRI, graphical models, group analysis, random walker.

1 Introduction

Functional magnetic resonance imaging (fMRI) has become one of the most widely-used modality for studying human brain activity. The standard approach for analyzing fMRI data involves separately comparing each voxel's intensity time course against an expected response to generate statistics that reflect the likelihood of activation [1]. The drawback to this univariate approach is that voxel interactions are ignored despite that each voxel is unlikely to function in isolation. To remedy this limitation, methods based on Markov random fields (MRF) [2] and Bayesian statistics [3] have been proposed to incorporate voxel interactions in the form of neighborhood information. These methods help suppress false declaration of isolated voxels as being active. However, the inherently low signal-to-noise (SNR) of fMRI data limits the reliability of the neighbors, which reduces the effectiveness of the currently-used regularization methods. The core of the problem is that there may just be insufficient information within a single subject's data to obtain satisfactory functional segmentation. Additional information is thus needed to disambiguate the state of noisy voxels.

Most fMRI studies focus on identifying common patterns across subjects, and thus exploiting the *group* dimension presents a direct, intuitive means of enhancing single-subject segmentations [4]. In standard fMRI group analysis, brain images of all

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 331–338, 2010.

subjects are first warped onto a common template to create a voxel correspondence [1]. Activation statistics are then compared across subjects to generate a group map. The underlying assumption is that a perfect one-to-one voxel correspondence is established after whole-brain warping. However, the vast anatomical variability renders this assumption questionable. In fact, even if perfect anatomical alignment is achieved, whether a one-to-one functional correspondence exists between voxels is debatable. Past studies have shown considerable functional inter-subject variability [5], which suggests that such one-to-one voxel correspondence is rather unlikely. However, active voxels are typically observed within the same anatomical regions across subjects [5]. Thus, integrating inter-subject neighborhood evidence is likely to help regularize single-subject segmentations and better distinguish signal from noise.

In this paper, we propose a new graph-theoretic method, "Group Random Walker" (GRW) that extends our previous work [6] on RW for estimating single-subject probabilistic activation maps. Treating each voxel as a graph vertex, we extend edges to inter-subject in addition to intra-subject neighboring voxels to jointly exploit group information and voxel interactions. Integrating group information into each subject's activation map, as opposed to estimating a group map, also facilitates inter-subject commonalities as well as differences to be modeled. Moreover, GRW draws upon a RW formulation [7] that provides an exact, unique closed-form solution for computing probabilistic activation maps with global optimality guaranteed.

2 Proposed Method

We propose extending the single-subject neighborhood system to other subjects within a group to disambiguate the state of noisy voxels. The intuition behind this approach is that true brain activation should appear in similar proximal locations across subjects [5], whereas false positives are more randomly scattered across the brain. Hence, regularizing inter-subject neighbors reinforces brain areas that are consistently recruited across subjects, while suppressing the false positives. Since only voxels that are spatially-proximal to each other are encouraged to be in similar state in our framework, the stringent one-to-one voxel correspondence requirement in standard fMRI group analysis is mitigated. Establishing an inter-subject neighborhood system requires first aligning the brain structures of all subjects. However, the vast anatomical variability renders accurate whole-brain warping difficult, especially for diseased subjects. Therefore, we instead employ a region-based approach, where we extract anatomical regions of interest (ROIs) and perform alignment at the regional level. This approach ensures that no brain structures will be mistakenly taken as part of another structure which has shown to improve activation localization [8].

2.1 Group Random Walker

In the original RW framework, each voxel is represented as a graph vertex with weighted edges added between spatial neighbors to bias the paths for which a random walker may transverse. Voxels are labeled (e.g. active or non-active) based on the probability that a random walker starting at each voxel location will first reach a pre-labeled seed. This framework, however, not only requires specifying seed voxels but

also does not model unary voxel information, such as activation effects in the context of fMRI. Therefore, we adopt an augmented RW formulation [7] that facilitates incorporation of unary information as label priors. This formulation is equivalent to adding an artificial seed vertex for each label and connecting these seeds to every vertex in the original graph with label priors being the edge weights [7]. The corresponding energy functional is as follows:

$$E(x^s) = x^{sT} L x^s + \sum_{k=1, k\neq s}^{K} x^{kT} \Lambda^k x^k + (x^s - 1)^T \Lambda^s (x^s - 1), \quad (1)$$

where x^s are the unknown posterior probabilities of the voxels belonging to label class s, K is the number of labels, Λ^s is a diagonal matrix containing prior probabilities of the voxels belonging to label class s (Section 2.2), and L is a weighted graph Laplacian matrix (Section 2.3). This construction is analogous to graph cuts, where the first term in (1) models voxel interactions, while the second term models unary voxel information. The main difference is that RW minimizes (1) over real-valued probabilities instead of binary numbers, which has an exact, unique closed-form solution with global optimality guaranteed for an arbitrary number of labels [7]. Specifically, x^s can be estimated by solving [7]:

$$\left(L + \sum_{k=1}^{K} \Lambda^k \right) x^s = \lambda^s, \quad (2)$$

where λ^s are the diagonal elements of Λ^s. To incorporate group information, we extend the augmented RW graph structure by inserting edges between each subject's voxels and their inter-subject neighbors. By treating all subjects' voxels as a single set and adding edges in the manner described later in Section 2.3, (2) can be directly applied to jointly estimate probabilistic activation maps of all subjects. GRW hence inherits all desired properties of the RW formulation. Globally optimal labeling can be obtained by assigning voxels to the labels associated with the highest probability.

2.2 Label Priors

To compute label priors λ^s, we first estimate the ROI activation statistics t_j of each subject using the standard general linear model (GLM) [1]:

$$y_j = X\beta_j + \omega_j, \; t_j = \beta_j / se(\beta_j), \quad (3)$$

where y_j is the time course of voxel j, ω_j is assumed to be white Gaussian noise after preprocessing, β_j are the estimated activation effects, and $se(\beta_j)$ is the standard error of β_j. X is a design matrix with boxcar functions (time-locked to stimulus) convolved with the hemodynamic response (HDR) as regressors [1]. We model t_j using a constrained Gaussian mixture model (CGMM) [9]. Specifically, t_j is assumed to be generated from a mixture of K Gaussian distributions with mixing coefficients π_k, means μ_k, and variance σ_k^2. Conjugate priors are used to constrain these parameters:

$$t_j \sim \sum_{k=1}^{K} \pi_k N(\mu_k, \sigma_k), \; \pi \sim Dir(\alpha), \; \mu_k \sim N(\eta, \tau^2), \; \sigma_k^2 \sim IG(a,b), \quad (4)$$

where $IG(a,b)$ and $Dir(\alpha)$ denote inverse Gamma and Dirichlet distributions. Adding priors enables us to integrate our knowledge into the model. In particular, we know that t_j of non-active voxels should theoretically be 0 and t-threshold for active voxels is typically set between 3 and 4 based on Gaussian Random Field (GRF) theory [1]. We thus encode this prior knowledge on t-values of active and non-active voxels through η with τ^2 set to 1 to model uncertainty in η. We set K to 2 to classify voxels as active or non-active [9]. As for σ_k^2, we use an uninformative prior by setting a and b to 0.5 [9], since little is known about σ_k^2. α is set to $1/K$ assuming equal prior class probabilities. Gibbs sampling is employed to estimate the probability of voxel j belonging to each of the K labels [9], which we use as label priors λ^s.

2.3 Weighted Graph Laplacian

Treating voxels of all subjects as a single set, we define the weighted graph Laplacian L based on functional connectivity $f_{ij} = correlation(Y_i,Y_j)$ and spatial distance d_{ij}:

$$L_{ij} = \begin{cases} -w_{ij}, & e_{ij} \in E_{\text{intra}} \cup E_{\text{inter}} \\ \sum_i w_{ij}, & i = j \\ 0, & \text{otherwise} \end{cases}, \tag{5}$$

where $w_{ij} = f_{ij}+exp(-d_{ij})$, Y_j is the magnitude spectrum of the Fourier transform of voxel j's time course y_j, e_{ij} denote a graph edge between voxels i and j, and E_{intra} and E_{inter} denote the sets of intra- and inter-subject edges. This choice of w_{ij} is motivated by the well-known bilateral filtering technique, which enables closer spatial neighbors with higher functional connectivity to exert greater influence on each voxel, thus de-weighting contributions from outliers. Edges are added between every given voxel of subject p and its 6-connected intra-subject spatial neighbors and c closest inter-subject spatial neighbors for every subject pairs (p,q), $p \neq q$. c is empirically set to 3. Note that voxel interactions are modeled using $correlation(Y_i,Y_j)$, instead of temporal correlations, since the temporal profile of HDR is known to vary across subjects [9]. In contrast, magnitude spectrums, Y_i, of active voxels would likely display higher similarity across subjects since all subjects are guided by the same stimulus.

2.4 Empirical Evaluation

500 synthetic datasets were generated to validate our proposed method. Each dataset consisted of 10 subjects with artificial activation injected within real, manually-segmented anatomical ROIs (Section 3). Voxels within a radius of 8 mm from the anatomical centroid were defined as active (circled in red in Fig. 1(b)-(f)). Synthetic time courses of the active voxels were generated by convolving a box-car function, having the same stimulus timing as our experiment (Section 3), with a canonical HDR [1] and adding low frequency drifts and Gaussian noise. To simulate functional inter-subject variability, signal intensity of the active voxels was set to decrease exponentially as a function of distance from the activation centroid, whose location was randomly varied across subjects (Fig. 1(a)). This emulates the situation where true active regions highly overlap across subjects, but the apparent overlap appears

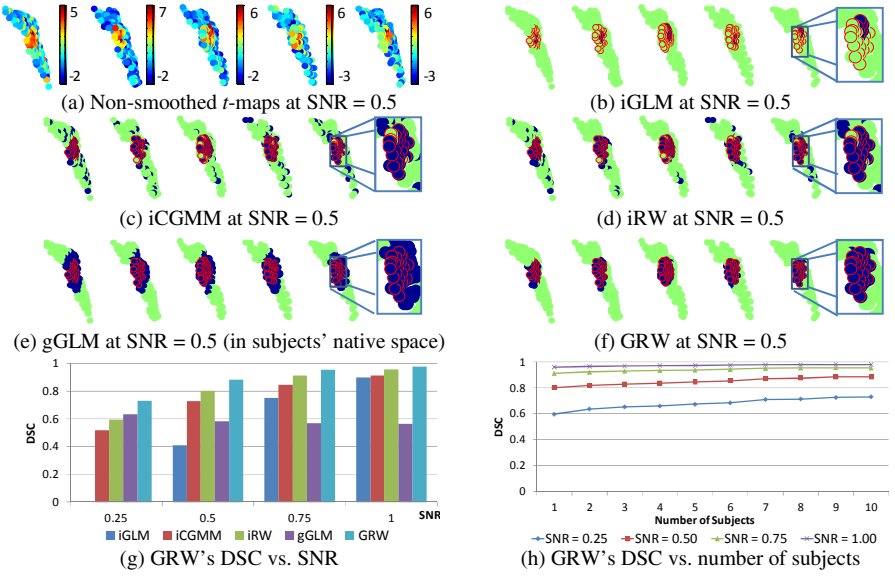


Fig. 1. Synthetic data results. (a) *t*-maps. (b) iGLM, (c) iCGMM, (d) iRW, (e) gGLM, and (f) GRW results shown. Blue dots = detected active voxels. Red circles = ground truth. For all SNR, (g) GRW achieved the highest DSC. Note iGLM resulted in a DSC of 0 at a SNR of 0.25. (h) GRW's DSC increased with number of subjects.

much less due to variation in location at which the fMRI signal concentrates. Maximal SNR was set as 0.5 in Fig. 1(a). For comparison, we also tested (i) GLM with spatial smoothing using a 8 mm FWHM Gaussian kernel and a threshold based on GRF theory for a p-value of 0.05 [1], (ii) CGMM, (iii) RW, and (iv) second level GLM, which involved taking the union of all subjects' ROI point sets to generate an ROI template, interpolating spatially smoothed β_j onto the template, applying GLM on the resulting β_j, thresholding based on GRF theory [1], and interpolating the thresholded group map back onto the subjects' native ROI space for comparison purposes. We refer to methods (i), (ii), (iii), and (iv) as individual GLM (iGLM), individual CGMM (iCGMM), individual RW (iRW), and group GLM (gGLM).

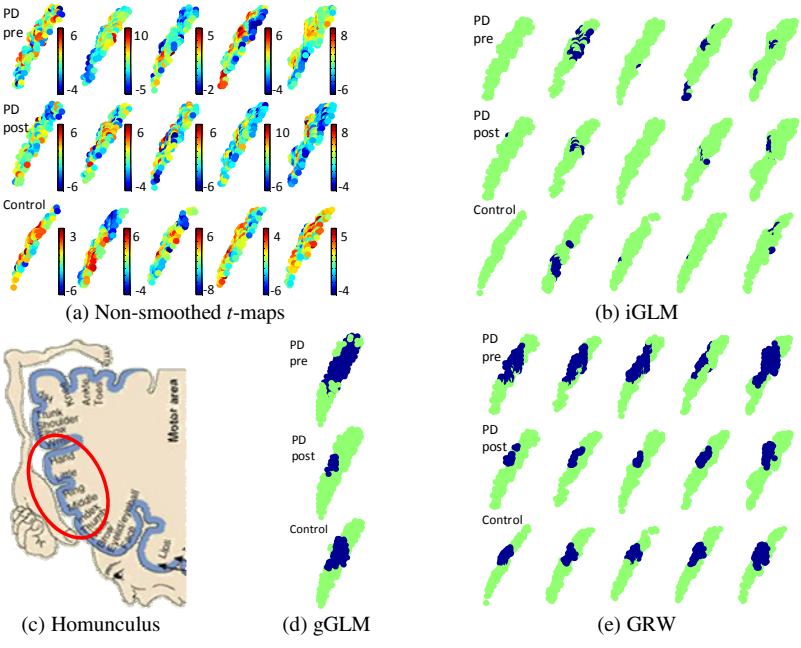
Qualitative results for the various contrasted methods are shown in Fig. 1(b)-(f). Only half of the subjects for one of the synthetic datasets are displayed due to space limitation. iGLM detected only a few active voxels, whereas iCGMM detected majority of the active voxels but also declared many false positives. Imposing intra-subject regularization using iRW reduced the number of false positives, but an ample amount remained due to lack of reliable intra-subject neighbors. Using gGLM detected all the active voxels, but also falsely declared many nearby voxels as active. Using GRW detected almost all the active voxels, while exerting much stricter control on false positives than gGLM. To quantify the performance, we computed the average Dice similarity coefficient (DSC) over the 500 synthetic datasets for a range of SNR.

$$\mathrm{DSC} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}, \tag{6}$$

where TP, FP, and FN denote the number of true positives, false positives, and false negatives, respectively. As evident from Fig. 1(g), the poor DSC for iGLM, iCGMM, and iRW again illustrates that solely relying on single-subject information may be inadequate to obtain satisfactory segmentation at low SNR. However, naively incorporating group information can also be problematic as apparent from the gGLM results, where increasing SNR reduced DSC. This counter-intuitive result arose from the increase in signals leaking into the non-active voxels as a consequence of spatial smoothing, as required in gGLM for employing GRF theory and increasing functional overlap across subjects. In contrast, increasing SNR resulted in higher DSC for GRW, since GRW does not blindly smooth the data. Instead, intra- and inter-subject neighborhood information is adaptively controlled based on functional connectivity with contributions from dissimilar voxels de-weighted. As a result, GRW achieved the highest DSC for all SNR compared to the other examined methods. Also, adding group information improved performance over using iGLM, iCGMM, and iRW even at higher SNR, where reliable intra-subject information is available. Furthermore, increasing the number of subjects increased DSC as shown in Fig. 1(h).

3 Materials

After obtaining informed consent, fMRI data were collected from 10 Parkinson's disease (PD) patients off and on medication (4 men, 6 women, mean age 66 ± 8 years) and 10 healthy controls (3 men, 7 women, mean age 57.4 ± 14 years). Each subject used their right hand to squeeze a bulb with sufficient pressure to maintain a bar shown on a screen within an undulating pathway. The pathway remained straight during baseline periods and became sinusoidal at a frequency of 0.25 Hz (slow), 0.5 Hz (medium) or 0.75 Hz (fast) during time of stimulus. Each session lasted 260 s, alternating between baseline and stimulus of 20 s duration. Functional MRI was performed on a Philips Gyroscan Intera 3.0 T scanner (Philips, Best, Netherlands) equipped with a head-coil. T2*-weighted images with BOLD contrast were acquired using an echo-planar (EPI) sequence with an echo time of 3.7 ms, a repetition time of 1985 ms, a flip angle of 90°, an in plane resolution of 128×128 pixels, and a pixel size of 1.9×1.9 mm. Each volume consisted of 36 axial slices of 3 mm thickness with a 1 mm gap. A T1-weighted image consisting of 170 axial slices was also acquired. For each subject's data, slice timing and motion correction were performed using Brain Voyager's (Brain Innovation B.V.). Further motion correction was then applied using motion corrected independent component analysis (MCICA) [10]. The voxel time courses were high-pass filtered to account for temporal drifts and temporally whitened using an autoregressive AR(1) model. No whole-brain warping or spatial smoothing was performed. For testing our proposed method, we selected the left primary motor cortex (LM1), which is known to activate during right-hand movements. Delineation of LM1 was performed by an expert based on anatomical landmarks and guided by a neurological atlas. The segmented ROIs were resliced at fMRI resolution for extracting preprocessed voxel time courses within each ROI and non-rigidly aligned using "Coherent Point Drift", which has shown greater robustness to noise and outliers than conventional techniques such as iterative closest point [11].



(a) Non-smoothed *t*-maps (b) iGLM

(c) Homunculus (d) gGLM (e) GRW

Fig. 2. Real data results for 5 PD subjects pre- and post-medication and 5 controls. Blue in (b), (d) & (e) indicates detected active voxels. (a) *t*-maps. (b) iGLM detected only a few active voxels in the (c) hand region, whereas (d) gGLM detected the hand region but falsely included the hip and leg areas for PDpre. (e) GRW correctly identified the hand region in all subjects.

4 Results and Discussion

Results obtained with iGLM, gGLM, and GRW on real data are shown in Fig. 2. iCGMM and iRW results were similar to the synthetic case with many isolated false positives detected, and were thus excluded. Also, only results for 5 controls and 5 PD subjects during the fast condition are displayed due to space limitation, but consistent results were observed across all subjects. iGLM detected few active voxels in the hand region, whereas gGLM detected the hand region, but mistakenly included the hip and leg areas for PD pre-medication. In contrast, GRW correctly identified the hand region in all subjects without falsely declaring the hip and leg areas as active. In addition, the GRW results suggest a very interesting trend across subject groups. Specifically, PD pre-medication seemed to require recruiting a wider area of LM1, which normalized back to an extent similar to the controls upon medication. Such spatial focusing effect of levo-dopa medication has been observed in past studies [12], thus further confirming the validity of our results. This trend is also noticeable in the gGLM results, but whether levo-dopa truly over-normalized the extent of activation in PD is unclear, since the wider active region in control subjects could have simply arisen from gGLM's weak control over false positives (Fig. 1(e) and Fig. 2(d) PD pre-medication). In contrast, the stronger control GRW has on false positives provides us more confidence with our findings, which greatly eases result interpretation.

5 Conclusion

We proposed a novel graph-theoretic method for enhancing single-subject fMRI activation detection. GRW expands the single-subject graph structure to include inter-subject neighbours, which enables group information to propagate into each subject's activation map without having to establish a one-to-one voxel correspondence. Also, the proposed GRW formulation permits joint estimation of all subjects' probabilistic activation maps with global optimality guaranteed. Superior detection power over standard techniques was shown on synthetic data for a range of SNR. When applied to real data, GRW consistently detected activation in regions implicated with the experimental task employed, whereas methods based on single-subject information failed. Our results thus demonstrate the effectiveness of incorporating group information for dealing with noisy fMRI data in single-subject analysis.

References

1. Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.B., Frith, C.D., Frackowiak, R.S.J.: Statistical Parametric Maps in Functional Imaging: A General Linear Approach. Hum. Brain Mapp. 2, 189–210 (1995)
2. Descombes, X., Kruggel, F., von Cramon, D.Y.: Spatio-Temporal fMRI Analysis Using Markov Random Fields. IEEE Trans. Med. Imaging. 17, 1028–1039 (1998)
3. Penny, W.D., Trujillo-Barreto, N.J., Friston, K.J.: Bayesian fMRI Time Series Analysis with Spatial Priors. NeuroImage 24, 350–362 (2005)
4. Bathula, D.R., Staib, L.H., Tagare, H.D., Papademetris, X., Schultz, R.T., Duncan, J.S.: Multi-Group Functional MRI Analysis Using Statistical Activation Priors. In: MICCAI fMRI Data Analysis Workshop (2009)
5. Vandenbroucke, M.W.G., Goekoop, R., Duschek, E.J.J., Netelenbos, J.C., Kuijer, J.P.A., Barkhof, F., Scheltens, P., Rombouts, S.A.R.B.: Interindividual Differences of Medial Temporal Lobe Activation during Encoding in an Elderly Population Studied by fMRI. NeuroImage 21, 173–180 (2004)
6. Ng, B., Abugharbieh, R., Hamarneh, G., McKeown, M.J.: Random Walker Based Estimation and Spatial Analysis of Probabilistic fMRI Activation Maps. In: MICCAI fMRI Data Analysis Workshop (2009)
7. Grady, L.: Multilabel Random Walker Image Segmentation Using Prior Models. In: IEEE Comp. Soc. Conf. Comp. Vision Pattern Recog., vol. 1, pp. 763–770 (2005)
8. Stark, C.E., Okado, Y.: Making Memories without Trying: Medial Temporal Lobe Activity Associated with Incidental Memory Formation during Recognition. J. Neurosci. 23, 6748–6753 (2003)
9. Makni, S., Ciuciu, P., Idier, J., Poline, J.B.: Joint Detection-Estimation of Brain Activity in Functional MRI: A Multichannel Deconvolution Solution. IEEE Trans. Sig. Processing. 53, 3488–3502 (2005)
10. Liao, R., Krolik, J.L., McKeown, M.J.: An Information-theoretic Criterion for Intrasubject Alignment of fMRI Time Series: Motion Corrected Independent Component Analysis. IEEE Trans. Med. Imaging. 24(1), 29–44 (2005)
11. Myronenko, A., Song, X., Carreira-Perpinan, M.: Non-rigid Point Set Registration: Coherent Point Drift. In: 20th NIPS, pp. 1009–1016. MIT Press, Cambridge (2006)
12. Monchi, O., Petrides, M., Doyon, J., Postuma, R.B., Worsley, K., Dagher, A.: The Neural Bases of Set-shifting Deficits in Parkinsons Disease. J. Neurosci. 24, 702–710 (2004)

Measures for Characterizing Directionality Specific Volume Changes in TBM of Brain Growth

Vidya Rajagopalan[1,2], Julia Scott[1,2], 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
{vidya.rajagopalan,julia.scott,piotr.habas}@ucsf.edu,
{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 7005 CNRS/University of Strasbourg, 67412 Illkirch, France

Abstract. Tensor based morphology (TBM) is a powerful approach to analyze local structural changes in brain anatomy. However, conventional scalar TBM methods are unable to present direction-specific analysis of volume changes required to model complex changes such as those during brain growth. In this paper, we describe novel TBM descriptors for studying direction-specific changes in a subject population which can be used in conjunction with scalar TBM to analyze local patterns in directionality of volume change during brain development. We illustrate the use of these methods by studying brain developmental patterns in fetuses. Results show that this approach detects early changes local growth that are related to the early stages of sulcal and gyral formation.

1 Introduction

Tensor based morphology (TBM) is now widely used as a method to detect structural brain differences across a population or across time. TBM studies involve statistical analysis of deformation fields computed from non-rigid registration of different anatomies. Scalar TBM, in its original form [1], is still being used to analyze structural changes caused by neurodegenerative conditions in adult brains [2]. But more recently, TBM is also being increasingly used for modeling of growth in developing brains [3] [4] [5]. However, a key weakness of scalar TBM is that it explicitly ignores directionality of volume changes. The multivariate, strain tensor metric proposed by Lepore et al. [6], captures anisotropic volume changes completely. However, this metric also ignores volume independent, local orientation changes within the brain, and is unable to specify the directionality of shape change for intuitive anatomical understanding [7]. Studying the complex shape changes during brain development is a key motivation for the extension of TBM to look at patterns of direction specific volume change. For example, development of the sylvian fissure in the fetal brain is related to superior-inferior expansions in the frontal and temporal lobes along with left-right contractions in the inferior aspect of the frontal cortex and superior aspect of the temporal cortex [8].

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 339–346, 2010.

In this paper, we propose a method that allows us to study both volumetric and directional changes associated with growth modeling. Specifically we use polar decomposition to extract local orientation information from the deformation tensors. This orientation information is represented in the form of a measure of common directionality of growth, which is then analyzed to detect regions in which the rate of growth in a direction significantly changes. In this paper we (i) propose a descriptor to quantify population-wide direction change patterns; (ii) combine these with conventional scalar TBM so as to study population-wide changes in both volume and direction; and (iii) present a test to extract the primary growth direction at each tissue location. We evaluate the proposed methods by applying them to a study of fetal brain growth.

2 Methods

For a cross-sectional population of N subjects, we compute a transformation $T_i \quad i = 1, 2, ...N$ for each subject, which maps the anatomical changes required to spatially normalize that particular subject to the average space. At each voxel p, the local changes can be derived from the deformation tensor which is defined as the gradient of the transformation $(T_i^p[x, y, z])$ at that voxel and is given by $(J_i^p) = \left[\frac{\delta^3 T_i^p[x,y,z]}{\delta x \delta y \delta z}\right]$

By computing J at each voxel, we form a map of local changes across the subject population. In this paper, we use the tensor decomposition properties of the Jacobian matrices to identify deformation direction.

2.1 Deformation Direction Vector (DDV)

The principal deformation direction can be obtained from the principal component of the Jacobian matrix at each voxel. In order to avoid stability issues associated with eigen analysis of any Jacobian matrix , we leverage the polar decomposition property of J [9] to compute the primary deformation direction indirectly. At each voxel, the Jacobian matrix, being a second order tensor, can be separated into a rotational and a strain component.

$$J = RS \tag{1}$$

where S is the symmetric, positive definite strain tensor ($S = (J^T J)^{\frac{1}{2}}$). $R = S^{-1}J$ is the orthogonal, rotation matrix such that $R^{-1} = R^T$ and $\det R = +1$. As illustrated Figure 1, this decomposition is analogous to first stretching the voxel by S and then rotating it by R to realize the deformation. Therefore, the direction of deformation, is computed by extracting the principal eigenvector of strain ($\mathbf{e}$) from S, and rotation this by R yields the DDV. Here, we would like to emphasize that $\mathbf{e}$ only specifies the principal direction of strain and not the direction of the complete deformation. For each subject, we compute a DDV map (DDV at each voxel) which is used to analyze local changes in directionality.

Statistical Analysis DDV. The DDV map can then be used in the TBM framework to detect population wide changes. In Equation 2, the basis vectors correspond to the three primary, orthogonal directions in reference anatomy: inferior-superior, left-right

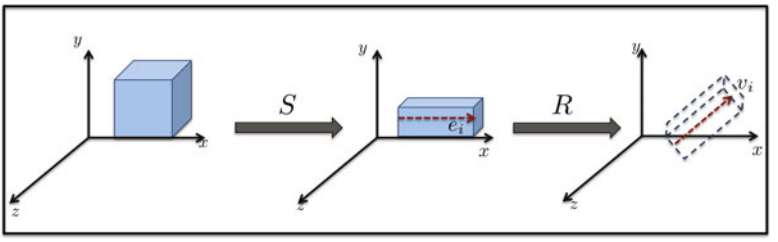


Fig. 1. Computation of deformation direction vector (DDV) from polar decomposition of Jacobian matrix

and anterior-posterior. The scalar multiplicative factors (x, y, z) from the DDV are separated to form three directional component maps (DCM). The DCM can be considered statistically independent within the single deformation tensor, as knowledge of the scalar component along one direction gives no information about the other two values in the DDV.

$$DDV\,(x_i, y_i, z_i) = x_i(1, 0, 0) + y_i(0, 1, 0) + z_i(0, 0, 1) \tag{2}$$

Regression: In order to minimize local variations in direction, the DCMs are individually smoothed by a Gaussian kernel. In this work, we chose the kernel size (σ= 2mm) experimentally so as to obtain the best final quality of the fitting. Using regression, we examine if there is a relationship between the basis vectors and variables of interest related to each subject (such as age or clinical criteria) using the multivariate general linear model at each voxel [10]. Growth models of DDV with respect to age can be performed using multivariate, multiple regression as shown in Equation 2.1. Here A corresponds to the vector of DCM values at each voxel, $V_1, \ldots, V_m$ are the independent variables (age, gender, clinical condition, etc.) and ε are the errors. (The matrix dimensions of each of the variables are indicated below each variables in Equation 2.1). Linear least square methods are used to solve for $\beta_1, \ldots, \beta_m$ at each voxel.

$$\underset{(n \times 3)}{A} = \underset{(n \times 1)(1 \times 3)}{V_1\beta_1} + \ldots + \underset{(n \times 1)}{V_m\beta_m + \varepsilon}$$

Hypothesis Testing: Resulting regression coefficient $(\beta_1, \ldots, \beta_m)$ maps are tested for significance using a standard t-test. Statistical significance was computed and these were corrected multiple comparisons using permutation tests [11]. The corrected p-value maps of the three directional components can then be analyzed individually or can be combined using Fisher's meta-analysis method [12] for independent tests. Let $h_o^1, h_o^2 \ldots h_o^k$ be the k independent null hypotheses being tested. Fisher's meta-analysis tests the null hypothesis that all the individual null hypotheses are true. The alternate hypothesis of the meta-analysis is that at least one of the individual null hypothesis is false. The p-values from the individual tests are combined using the formula $\kappa = -2\sum_{i=1}^{k} \log_e(p_i)$ where p_i is the p-value of the i^{th} individual hypothesis test. The meta-analysis statistic κ has a chi-squared distribution under the null hypothesis with $2k$ degrees of freedom.

The $\beta_1, \ldots, \beta_m$ values are estimates of increase or decrease in volume along a particular direction and the hypothesis tests estimate statistical significance of these changes. Multivariate hypothesis tests are not well suited for this metric as a meaningful covariance between directions cannot be established.

2.2 Principal Growth Direction (PGD)

In scalar TBM, the definition of volume change is well understood i.e. undeformed voxels have a Jacobian determinant of 1 and any deviation from this value is considered as change. For directional properties such a distinct baseline does not exist. Therefore, here we propose to start by first computing the most common growth direction at each tissue location across the population being studied and then analyze how directional growth is varying with respect to this direction at each location. For this we make use of the DDV map for each subject which provides orientation information specific to each subject at any given location. As described in Equation 3, PGD at a voxel is defined as the DDV (of a single subject) which has the least circular distance [13] from DDVs of all the other $(N-1)$ subjects.

$$PGD = \arg\min_{DDV_i} \left(\sum_{j=1}^{N} \frac{1}{2} \left(1 - \frac{DDV_i \cdot DDV_j}{|DDV_i||DDV_j|} \right) \right); \quad \forall i \neq j;\ i = 1, 2, \ldots N \quad (3)$$

The PGD map gives us a baseline which can be used along with the results from the directional TBM in Section 2.1 to study how these directionality of growth changes with age and development[1].

3 Application – Early Fetal Brain Growth

The following experiments were performed using clinical MR scans of 40 fetal subjects at gestational ages ranging from 20.57 to 27.86 weeks. The mothers were referred for fetal MRI due to questionable abnormalities on prenatal ultrasound or a prior abnormal pregnancy. All women had normal fetal MRI and all newborns have had normal postnatal neurodevelopment. Fetal imaging was performed in our institution on a 1.5T scanner (GE Healthcare, Mulwaukee, WI) without sedation of the mother or the fetus. For each subject multiple stacks of single-shot fast spin-echo (SSFSE) T2-weighted slice images (pixel size 1 mm $\times$ 1 mm, slice thickness $\approx$ 3 mm) re acquired in the approximately axial, sagittal and coronal planes with respect to the fetal brain. The MR sequence parameters (TR = 4500 ms, TE = 91 ms) were originally designed for clinical scans and cannot be adjusted for image analysis purposes. High resolution 3D volumes were reconstructed from 2D slice MR images using the slice intersection motion correction (SIMC) technique [14]. The reconstructed volumes were automatically segmented into

[1] If a PGD cannot be clearly established i.e. the eigenvalues are not well-separated at many locations, the study can be done on each of the 3 directions separatly. 3 deformation direction maps are formed and are studied individually. The 3 significance maps are combined using an OR method or the fisher test.

individual tissue types (developing grey matter, developing white matter, the germinal matrix) using an atlas-based approach with probabilistic atlases generated from a spatiotemporal model of the fetal brain [15]. Smoothed tissue segmentation atlases were then used with a template free elastic registration method [16] to spatially normalize the subjects. For each subject, the Jacobian matrix maps were computed, from the resulting deformation fields, to quantify the pattern of deformation required to spatially normalize individual anatomies. For each subject, we computed the DDV maps and performed multivariate, multiple linear regression on the population with age as the independent variable and the directional coefficients as the dependent variables. The regression coefficients were tested for statistical significance. We also computed, using the DDV maps, a PGD map for the subject population.

3.1 Results and Discussion

Figure 2 shows the principal growth direction for the given age range overlaid on the spatially normalized average MRI, and displayed using the Rview software[2]. Overall growth shows a distinct spatial pattern. The intermediate zone (developing white matter (WM)), along the most dorsal and ventral areas, is growing primarily in the anterior-posterior (A-P) direction during this period. Along the cortical plate, PGD alternates between superior-inferior (S-I) and right-left (R-L) growth direction corresponding to the formation of sulci and gyri respectively. At the sylvian fissure we see that the growth is primarily along the R-L direction corresponding to the deepening of the fissure. The ventricles do not show a clear direction of growth as during this period in fetal brain development, the ventricles do not change significantly in absolute size as the brain grows.

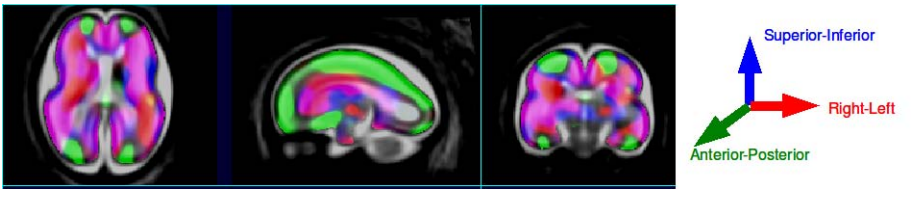


Fig. 2. Principal growth direction (PGD) vectors for the group of fetal anatomies studied. Each reference anatomical direction is indicated by one color: red = left-right (R-L); green = anterior-posterior (A-P); blue = superior-inferior (S-I). The panels from left to right indicate axial, sagittal and coronal views of the brain.

Figure 3 shows the regions where directional growth has changed significantly with age. The regions showing significant changes correspond to regions of major cortical folding. The largest cluster of significant voxels occur at the sylvian fissure where the significant acceleration of growth in the R-L direction and deceleration along the S-I direction indicate deepening of the fissure. As indicated by the cross-hairs, we see that the in addition to growth acceleration in the R-L direction, there is also a A-P growth

[2] http://rview.colin-studholme.net

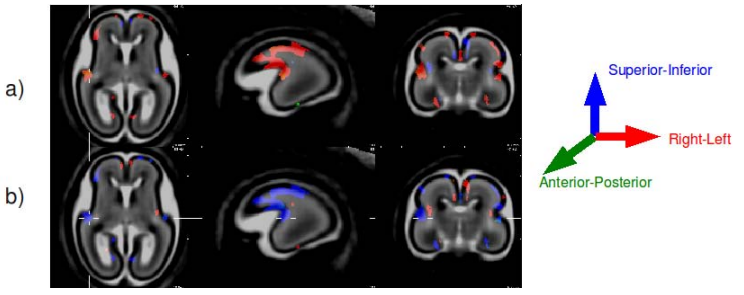


Fig. 3. Growth rate maps showing regions exhibiting significant change in growth direction with age. Row (a): Regions showing accelerated growth the associated anatomical direction. Row (b): Regions showing decelerated growth and the associated anatomical direction. Each reference anatomical direction is indicated by one color: red = left-right (R-L); green = anterior-posterior (A-P); blue = superior-inferior (S-I). The panels from left to right indicate axial, sagittal and coronal views of the brain.

component indicating "flattening" of the superior aspect of the temporal lobe at the sylvian fissure. To accommodate these changes in the cortical regions, we see that the intermediate zone underlying the fissure is being "stretched" in the S-I direction its growth in the R-L direction is restricted by neighboring structures.

Considering both Figures 2 and 3 together, we see that brain growth is characterized by spatially varying, directional growth. In some regions this directionality does not change significantly as the fetus matures. For example, the intermediate zone which showed a very strong A-P growth direction in Figure 2 does not exhibit significant accelerations and decelerations in any direction in the period of growth considered for this study. Major shape changes in the cortex occur due to significant changes in directional growth at sites of sulci and gyri. The effect of these changes in direction on underlying tissue is determined by the rate of those changes in the given time-period. For example, the rate of directional change associated with the formation of the calcrine sulcus is not large enough to significantly change directions in the underlying WM. In comparison,

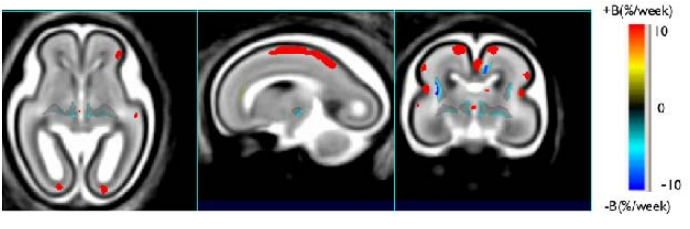


Fig. 4. For comparison we have included growth rate (specified by regression coefficients, B) maps computed using scalar TBM. B value maps are overlaid on average MR image. Yellow and red regions correspond to accelerated growth and shaded blue regions correspond to decelerated growth when compared to supratentorial brain growth.

the formation of the sylvian fissure occurs at a more rapid rate and thereby significantly changes growth directionality both in the cortex and the underlying WM.

4 Conclusion

In this paper, we have introduced two novel descriptors that quantify directionality of change. This is the first time that local rotational information obtained from the deformation vector is being included in TBM analysis. The two descriptors are complementary in that the PGD describes the similarities in growth directions among the subject population and DDV regression allows us to detect changes in growth direction. The DDV can be incorporated into standard TBM framework and can be used along with other descriptors of volumetric changes. The clinical value of this method increases since we are able to specify a single direction of growth for the sake of intuitive understanding. The inclusion of directional growth information with standard TBM, allows us to detect additional changes in tissue structure. Directional information is particularly helpful in the study of brain development. Here, the combination of volume and direction change patterns are able to better explain the mechanism of brain growth than the study of volumes changes alone. That being said, the use of these descriptors need not be restricted to only brain development studies. Any TBM application where brain structural differences (within or between groups) arise from directional variations in tissue gain/loss can be studied using the proposed descriptors.

Acknowledgments. This research was funded by NIH/NINDS grants: R01 NS 061957 and R01 NS 055064. Imaging for this study was also partially supported by the National Institutes of Health (NIH) Grant No. K23 NS52506-03 and NIH/NCRR UCSF-CTSI Grant No. UL1 RR024131. The work of F. Rousseau was supported by European Research Council under FP7/2007-2013 Grant Agreement 207667.

References

1. Davatzikos, C., Vaillant, M., Resnick, S., Prince, J., Letovsky, S., Bryan, R.: A computerized approach for morphological analysis of the corpus callosum. J. Comput. Assist. Tomogr. 20(1), 88–97 (1996)
2. Browndyke, J.N., Virk, H., Tupler, L.A., Warren, L.H., Hayden, K.M., Burke, J.R., Welsh-Bohmer, K.A., Doraiswamy, P.M.: Left anterior parahippocampal gray matter volume is inversely associated with APOE genotypic risk in healthy seniors: tensor-based morphometry data. Alzheimer's and Dementia 5(4, suppl. 1), P211–P212 (2009)
3. Lenroot, R.K., Gogtay, N., Greenstein, D.K., Wells, E.M., Wallace, G.L., Clasen, L.S., Blumenthal, J.D., Lerch, J., Zijdenbos, A.P., Evans, A.C., Thompson, P.M., Giedd, J.N.: Sexual dimorphism of brain developmental trajectories during childhood and adolescence. NeuroImage 36(4), 1065–1073 (2007)
4. Lee, J., Fonov, V., Evans, A.: Mapping brain growth of early childhood using deformation based morphometry. NeuroImage 47(suppl. 1), S153 (2009)
5. Gogtay, N., Giedd, J., Lusk, L., Hayashi, K., Greenstein, D., Vaituzis, A., Nugent, T., Herman, D., Clasen, L., Toga, A., Rapoport, J., Thompson, P.: Dynamic mapping of human cortical development during childhood through early adulthood. Proc. Natl. Acad. Sci. USA 101(21), 8174–8179 (2004)

6. Lepore, N., Brun, C.A., Chiang, M.C., Chou, Y.Y., Dutton, R.A., Hayashi, K.M., Lopez, O.L., Aizenstein, H., 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)
7. Zhang, H., Awate, S., Das, S., Woo, J., Melhem, E., Gee, J., Yushkevich, P.: A tract-specific framework for white matter morphometry combining macroscopic and microscopic tract features. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 141–149. Springer, Heidelberg (2009)
8. Chi, J., Dooling, E., Giles, F.: Gyral development of human brain. Annals of Neurology 1(1), 86–93 (1977)
9. Golub, G., Van Loan, C.: Matrix Computations, 3rd edn. Johns Hopkins Studies in Mathematical Sciences. The Johns Hopkins University Press, Baltimore (1996)
10. Studholme, C., Cardenas, V.: Population based analysis of directional information in serial deformation tensor morphometry. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 311–318. Springer, Heidelberg (2007)
11. Nichols, T., Holmes, A.: Nonparametric permutation tests for functional neuroimaging: A primer with examples. Human Brain Mapping 15, 1–25 (2002)
12. Fisher, R.: Statistical Methods for Research Workers. Oliver and Boyd (1932)
13. Lund, U.: Least circular distance regression for directional data. Journal of Applied Statistics 26(6), 723–733 (1999)
14. Kim, K., Habas, P., Rousseau, F., Glenn, O., Barkovich, J., Studholme, C.: Intersection based motion correction of multislice MRI for 3-D in utero fetal brain image formation. IEEE Transactions on Medical Imaging 29(1), 146–158 (2010)
15. Habas, P., Kim, K., Rousseau, F., Glenn, O., Barkovich, J., Studholme, C.: A spatio-temporal atlas of the human fetal brain with application to tissue segmentation. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 289–296. Springer, Heidelberg (2009)
16. Studholme, C., Cardenas, V.: A template free approach to volumetric spatial normalization of brain anatomy. Pattern Recogn. Lett. 25(10), 1191–1202 (2004)

Inter-subject Connectivity-Based Parcellation of a Patch of Cerebral Cortex

Pauline Roca[1,2], Alan Tucholka[1,2,3], 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
[3] INRIA Saclay, Parietal, France

Abstract. This paper presents a connectivity-based parcellation of the human post-central gyrus, at the level of the group of subjects. The dimension of the clustering problem is reduced using a set of cortical regions of interest determined at the inter-subject level using a surface-based coordinate system, and representing the regions with a strong connection to the post-central gyrus. This process allows a clustering based on criteria which are more reproducible across subjects than in an intra-subject approach. We obtained parcels relatively stable in localisation across subjects as well as homogenous and well-separated to each other in terms of connectivity profiles. To address the parcellation at the inter-subject level provides a direct matching between parcels across subjects. In addition, this method allows the identification of subject-specific parcels. This property could be useful for the study of pathologies.

Keywords: human connectome, anatomical connectivity, cortical parcellation.

1 Introduction

In-vivo parcellation of the human cortex into functional brain areas is a major goal in neurosciences and clinical surgery. Anatomical connectivity based on diffusion-weighted imaging has been used to address this problem, based on the hypothesis that functional regions have a specific connectional fingerprint [1]. It presents the advantage of providing optimal basic elements for the construction of *the human connectome* [2]. These elements are defined at a scale maximizing connectivity-based similarities across subjects. However the huge dimension of connectivity data leads to important difficulties. A first approach consists in performing clustering of a small area of the brain using cross-correlation between connectivity profiles towards the full brain [3]. However, the huge dimension of the data has prevented from performing the clustering at inter-subjects level as well as on the whole brain cortex.

To overcome some of the limitations, a key idea is to collapse the connectivity profiles using a set of Regions Of Interest: connectivity weights are summed up across each ROI. Some approaches use a priori anatomical information such as

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 347–354, 2010.

lobar or gyri cortex segmentation [4,5] or regions of interest from invasive tracing primate studies [6]. When there is a correspondence across subjects between these segmentations, it provides a direct way to perform the clustering at the level of the group of subjects [4,5]. The robustness is largely increased because the clustering is focusing on profiles reproducible across subjects. Nevertheless using a cortex parcellation to infer another presents an evident bias. Another approach [7] avoids this by infering automatically the required segmentation from the connectivity data, independently for each subject. However this method has shown poor reproducibility across subjects.

The main contribution of this paper is to push this last approach at the level of the group of subjects. For this purpose, the set of ROIs obtained from the connectivity data is defined directly at the group level. This is achieved using a 2D coordinate system providing correspondence between the cortical surfaces across subjects [8]. Once the profiles have been collapsed using this set of ROIs, they are clustered following two different stategies. The profiles are either gathered alltogether or averaged across subjects. In the following the whole method is applied to parcellate the right post-central gyrus of a group of ten subjects.

2 Material and Methods

2.1 Data and Pre-processings

The present study has been performed on $N = 10$ subjects of the NMR database [9]. Diffusion weighted data were acquired with HARDI scheme, resulting in high-quality datasets based on 200 directions of diffusion and a b value of 3000 s/mm^2. Furthermore, the alignment between diffusion data and T1-weighted images is achieve by a dedicated MR sequence and distortion corrections. In order to recover the cortico-cortical anatomical connectivity, a streamline probabilistic tractography [10] is performed in the white matter volume using a field of Orientation Distribution Functions (ODF) calculated from the DW data with the analytical QBall model described in [11].

The grey/white matter interface is reconstructed using Freesurfer [12]. We created a slightly inflated inter-subject average mesh for visualisation purpose [8]. A point-to-point correspondence between all the meshes provides a way to analyze the connectivity data in a common space. The post-central gyrus is determined from the automatic surface labeling process of Freesurfer, which parcellates the surface into 66 anatomical cortical regions [13].

The cortical connectivity matrix of the post-central gyrus of each subject is calculated using the previous tractography results and surfacic data [7]. For each subject we obtained a matrix of size (G, C) where C and G are the number of vertices of the *cortex mesh* and of the *post-central gyrus* respectively. Each line of these matrices corresponds to the *connectivity profile* from one post-central gyrus vertex of one subject towards the full cortex defined in Freesurfer coordinate system. In order to take into account the uncertainty of the tractography,

a Gaussian smoothing of the connectivity matrix was applied on the surface ($\sigma = 5\ mm$).

2.2 Inter-subject Dimension Reduction of the Connectivity Data

In this part the objective is to compress the connectivity profiles in a way maximizing similarities across subjects. The sources of variability of connectivity profiles across subjects are multiple:

1. the artifactual variability created by weaknesses of the diffusion-based tracking scheme;
2. the non perfect spatial normalization;
3. the purely anatomical variability acting on areas of the parcels and on density and repartition of the fibers inside each parcel.

Geodesic smoothing of connectivity information is a first way to overcome the resulting difficulties. A second complementary way consists in performing a kind of group analysis detecting the cortical regions with reproducible connectivity to the input patch (here post-central gyrus). For each subject the post-central *gyrus connectivity profile* is computed, that is to say the average of the connectivity profiles across gyrus vertices (Fig. 1). This gyrus profile is normalized by the number of tracts reaching the gyrus. Then an average of these profiles across subjects is performed. The complete surface is masked to keep only nodes that received at least ten fiber tracts from at least half of the subjects.

The resulting *average connectivity profile* is then mapped on the average mesh(Fig. 1), represented as a texture. A watershed is computed for the *average connectivity profile* texture in order to split the cortical surface into catchment basins [7]. Each basin catches a set of tracts supposed to connect the post-central gyrus with a specific brain area (Fig. 1). This idea justifies the use of this set of basins to collapse the profiles in an optimal way regarding connectivity. Indeed, because of the inter-subject variability, the regions with a strong connection to the input patch (appearing in the connectivity profile of each

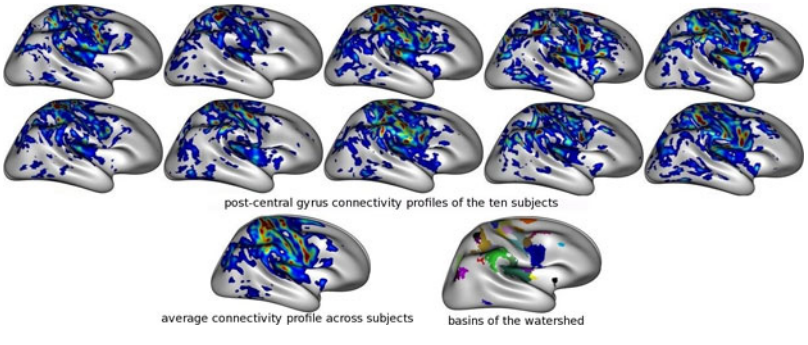


Fig. 1. Different connectivity profiles and associated watershed basins

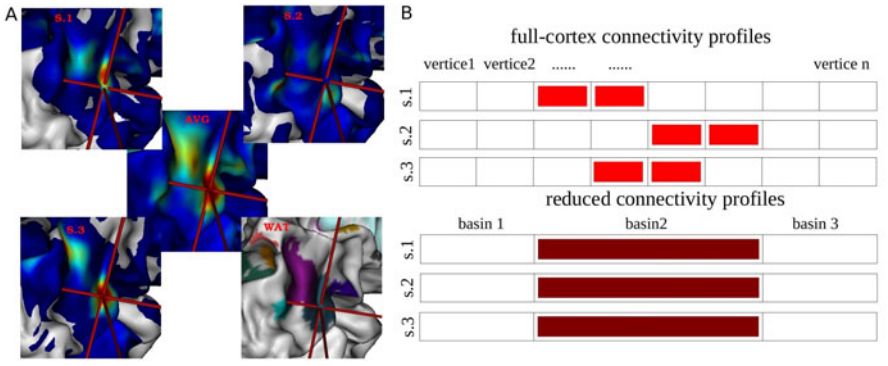


Fig. 2. Contribution of the watershed. A. *s1,s2,s3*: Zoom on one part of the post-central gyrus connectivity profile for three subjects; *avg.* Averaged connectivity profile; *wat.* Associated watershed basins. B. The same illustration in 1D.

subject) do not match perfectly when they are mapped onto the average cortex mesh(Fig.2.A.s1.s2.s3). When these profiles are averaged across subjects, the previous regions are grouped together to form a unique region (Fig.2.A.avg). The watershed algorithm can then isolate this region as a basin (Fig.2.A.wat). Note that a more sophisticated fMRI-inspired group analysis could be performed.

To conclude this step, these basins are used to reduce the dimension of the connectivity profiles: for each subject independently, all the vertices connectivity profiles (G profiles of size C) are collapsed to get G connectivity profiles of dimension n_{basins} (number of basins). This dimension reduction cancels out most of the inter-individual variability mentioned above. The connectivity information resulting from each brain area represented by a catchment basin is summarized to one single weight (Fig.2.B).

2.3 Cortical Parcellation

The reduced connectivity profiles of all the subjects are normalized using L_2-norm and then clustered using two alternative strategies.

First, the post-central gyrus vertices connectivity profiles of all the subjects (Fig.3.A) are concatenated into a big matrix of size $(N * G, n_{basins})$ (Fig.3.B). Hence the clustering algorithm deals with all the profiles without knowledge of the underlying subjects. The clustering is performed with the classical kmedoids algorithm (called PAM in R language [14]) and the Euclidean distance between profiles as dissimilarity measure. The number of clusters superior to two and maximizing the average silhouette width of the clustering is chosen. A cluster usually includes profiles stemming from most of the subjects. Therefore, the end result is one specific parcellation for each subject, but with a direct matching between the parcels across the subjects.

In the second approach, the same connectivity profiles are averaged across subjects for each vertice of the post-central gyrus (Fig.3.C). Then they are

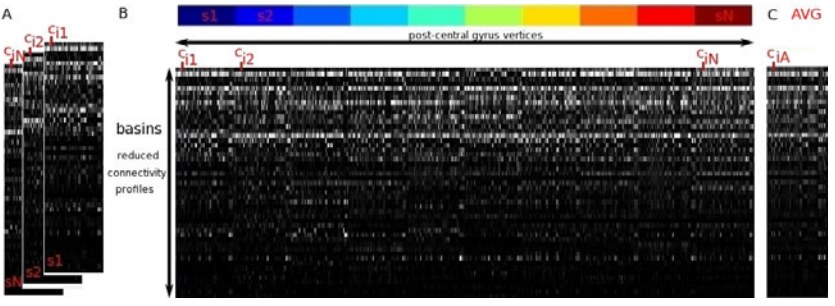


Fig. 3. Connectivity profiles of gyrus vertices: A. By subject; B. Concatenated for all the subjects; C. Averaged across subjects ($c_{iA} = \frac{1}{N}\sum_{s=1}^{N} c_{is}$)

clustered using the same algorithm (PAM). By doing this, one single parcellation is obtained from a connectivity matrix of relatively small size (G, n_{basins}).

The first approach has the advantage to preserve the inter-subject variability. Indeed the surface-based coordinate system is just used to define the input patch and to regroup the vertices into watershed basins for dimension reduction. The information related to the node localization in the coordinate system is not used by the clustering. When averaging, individual information is canceled out to provide a group result.

3 Results

The results are quite encouraging. The parcellations obtained by the two previous approaches are consistent with each other. For the average subject clustering, the optimal number of clusters is five (Fig.4) while with the concatenation approach, the optimal number is seven (Fig.5). For the concatenation approach, parcels have their own connectivity pattern, which is quite similar across subjects and relatively well-separated from other parcels (Fig.5.B). Note that in

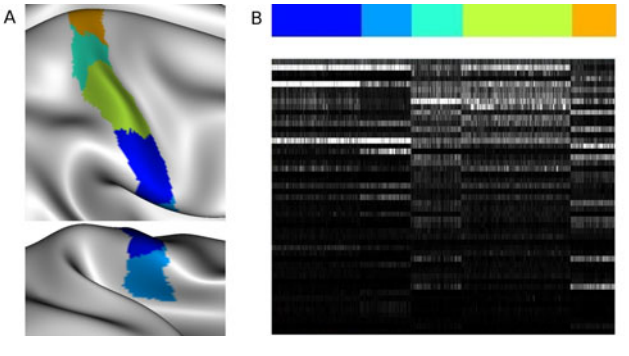


Fig. 4. Average approach: A. Parcellation; B. Averaged connectivity profiles (like in Fig.3.C) ordered by parcel

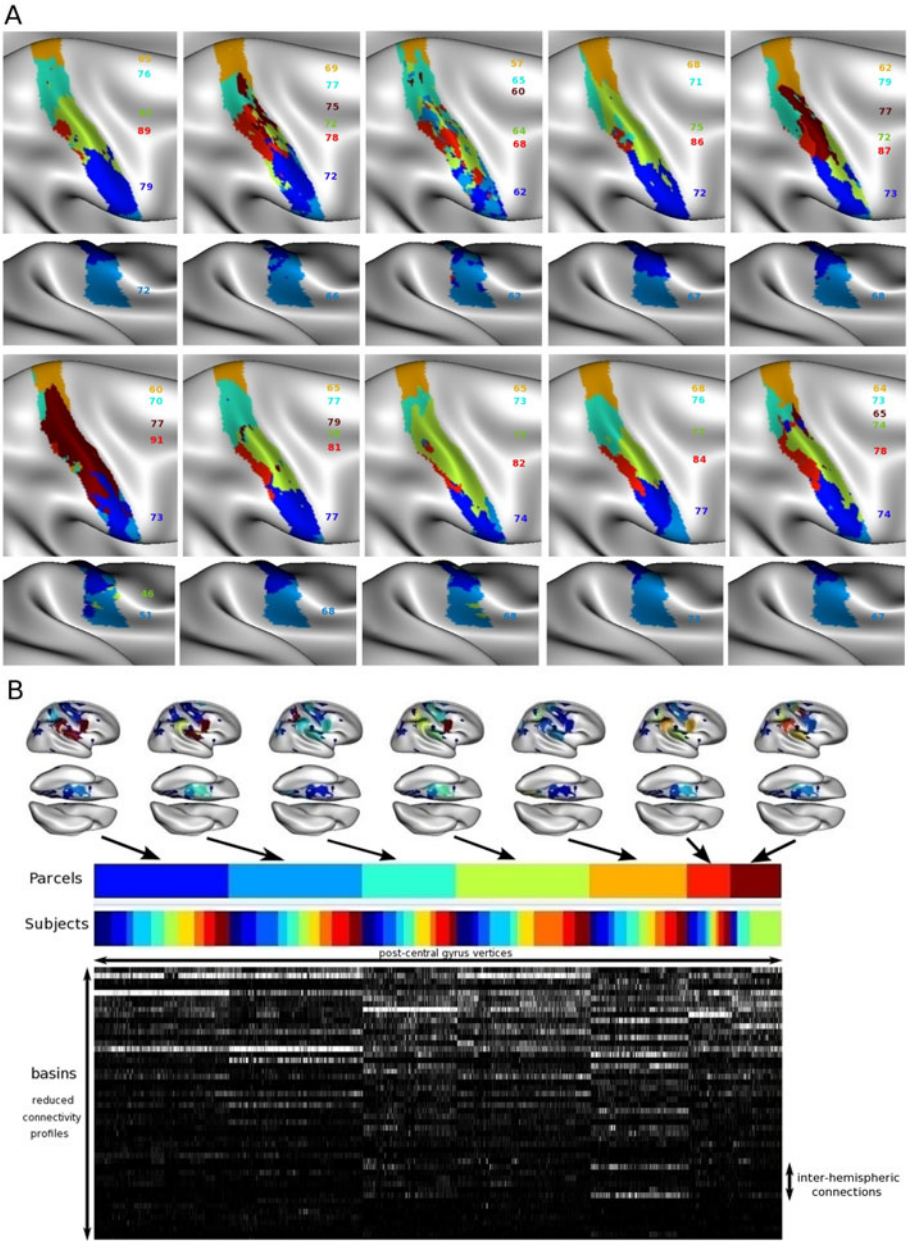


Fig. 5. Concatenation Approach: A. Parcellation for the ten subjects with the amount of tracts used by the clustering, by subject and parcel in percentage; B. i.Connectivity profiles of each parcel averaged across subjects and mapped on the inflated average mesh; ii. Concatenated connectivity profiles (like Fig.3.B) ordered by parcel

the concatenation approach, while the global number of parcels is seven, some parcels exist only for a subset of subjects. With a set of subjects composed of distinct sub-groups (patients vs controls for example) this flexibility could allow the identification of group-specific parcels.

The correspondence between the parcels found with the two approaches is straightforward considering the connectivity profiles and the localization (cf. parcels colors matching between Fig.4 and 5). One of the additional clusters for the concatenation approach stems mainly from two subjects (claret color in 5). It could result from artifacts in the underlying diffusion data leading to a spurious cluster. The second additional cluster (red) results from a split of the green cluster into two different parcels. This subdivision can not be detected by the average approach because of the interindividual variability of the localization of the boundary between green and red parcels. Except the claret parcel, the results of the concatenation approach are quite reproducible across subjects in terms of connectivity profiles and localization on the cortex. Note that the topology of the set of parcels is consistent across subjects and corresponds to the topology obtained by the average approach. The shapes and areas of each parcel and the localization of the boundaries between parcels, however, vary largely between subjects, which is consistent with literature on architectony.

4 Discussion

In this paper, we have shown that addressing the connectivity-based parcellation problem at an inter-subject level can lead to reproducible results even without using a priori knowledge for dimension reduction. In the optimal case, the watershed onto the average connectivity profile across subjects allows to regroup the subject-specific connections into one ROI used later to reduce the connectivity dimension. When it is not the case, the application of fMRI-inspired group analysis could provide better results and deserves our attention in the future. When dealing with a very large set of brains and targetting the parcellation of the complete cortical surface, the average approach may be the only reasonable way because of computational considerations. Note that extending the parcellation algorithm to the complete surface can be done iteratively like in [7]. On the other hand, when the concatenation approach is affordable, it provides a very attractive tool. Indeed, the connectome nodes are defined at the level of each subject, providing a connectivity-based referential to be used for various applications. For instance this parcellation could be the basis for fMRI or cortical thickness studies. The connectivity-based parcellation, indeed, provides an architectural spatial normalization of greater value than the traditional normalization based on the cortical surface geometry. Adapting the connectome to each subject connectivity data is also of great interest to push further the study of its topology using graph theory [15]. Defining the connectome nodes from the standard geometry based normalization indeed, is bound to hide the fine scale structure of the brain connectivity. In addition, the concatenation approach allows the identification of group-specific parcels and could be useful for the study of pathologies.

References

1. Passingham, R.E., Stephan, K.E., Kötter, R.: The anatomical basis of functional localization in the cortex. Nature 3, 606–616 (2002)
2. Sporns, O., Tononi, G., Kötter, R.: The human connectome: A structural description of the human brain. PLOS Computational Biology 1, 245–251 (2005)
3. Anwander, A., Tittgemeyer, M., von Cramon, D., et al.: Connectivity-based parcellation of brocas area. Cerebral Cortex 17, 816–825 (2007)
4. 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)
5. Jbabdi, S., Woolrich, M.W., Behrens, T.E.J.: Multiple-subjects connectivity-based parcellation using hierarchical dirichlet process mixture models. NeuroImage 44, 373–384 (2009)
6. Rushworth, M.F.S., Behrens, T.E.J., Johansen-Berg, H.: Connections patterns distinguish 3 regions of human parietal cortex. Cerebral Cortex 16, 1418–1430 (2005)
7. Roca, P., Riviere, D., Guevara, P.: Tractography-based parcellation of the cortex using a spatially-informed dimension reduction of the connectivity matrix. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 935–942. Springer, Heidelberg (2009)
8. Fischl, B., Sereno, M., Tootell, R., et al.: High-resolution intersubject averaging and a coordinate system for the cortical surface. Human Brain Mapping 8, 272–284 (1999)
9. Poupon, C., Poupon, F., Allirol, L., et al.: A database dedicated to anatomo-functional study of human brain connectivity. In: HBM 2006 (2006)
10. Perrin, M., Cointepas, Y., Cachia, A., Poupon, C.: Connectivity-based parcellation of the cortical mantle using q-ball diffusion imaging. Int. J. Biomed. Imaging (2008)
11. Descoteaux, M., Angelino, E., Fitzgibbons, S., et al.: Regularized, fast, and robust analytical q-ball imaging. Magnetic Resonance in Medicine 58, 497–510 (2007)
12. Dale, A.M., Fischl, B., Serenot, M.I.: Cortical surface-based analysis. i. segmentation and surface reconstruction. NeuroImage 9, 179–194 (1999)
13. Desikan, R., Ségonne, F., Fischl, B., et al.: An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. Neuroimage 31, 968–980 (2006)
14. Kaufmann, L., Rousseeuw, P.J.: Finding groups in data: an introduction to cluster analysis. Wiley Interscience, Hoboken (1990)
15. Hagmann, P., Cammoun, L., Gigandet, X., et al.: Mapping the structural core of human cerebral cortex. PLOS Computational Biology 6(7), 1479–1493 (2008)

On Super-Resolution for Fetal Brain MRI

F. Rousseau[1], K. Kim[2], C. Studholme[2], M. Koob[3], and J.-L. Dietemann[3]

[1] LSIIT, UMR 7005, CNRS - Université de Strasbourg, Strasbourg
[2] Biomedical Image Computing Group, Department of Radiology and Biomedical Imaging, University of California San Francisco
[3] LINC, FRE 3289, CNRS - Université de Strasbourg, Strasbourg

Abstract. Super-resolution techniques provide a route to studying fine scale anatomical detail using multiple lower resolution acquisitions. In particular, techniques that do not depend on regular sampling can be used in medical imaging situations where imaging time and resolution are limited by subject motion. We investigate in this work the use of a super-resolution technique for anisotropic fetal brain MR data reconstruction without modifying the data acquisition protocol. The approach, which consists of iterative motion correction and high resolution image estimation, is compared with a previously used scattered data interpolation-based reconstruction method. To optimize acquisition time, an evaluation of the influence of the number of input images and image noise is also performed. Evaluation on simulated MR images and real data show significant improvements in performance provided by the super-resolution approach.

1 Introduction

Imaging moving subjects remains an open issue for Magnetic Resonance Imaging (MRI). Although the development of ultrafast 2D acquisition sequences has led to significant improvements for clinical studies (see for instance [1] for fetal studies), the slice acquisition time is still critical and has to be as short as possible to reduce the impact of the motion. As a result, sets of thick 2D slices are generally acquired in clinical studies and interpretation remains limited by visual inspection.

Several clinical imaging protocols make use of multiple orthogonal 2D multiplanar acquisitions with non-isotropic voxel size for brain studies. In the context of fetal imaging[1], Rousseau *et al.* in [4] have proposed a registration-based method to compound multiple orthogonal sets of 2D fetal MRI slices into a single isotropic high resolution (HR) volume. This algorithm has been shown to be capable of reconstructing 3D geometrically consistent images from challenging clinically acquired data. The image reconstruction process is based on a local

[1] Please note that the problem of MR image reconstruction from motion corrupted data is not specific to fetal or neonatal MR imaging. It can occur also when imaging ageing patients or patients suffering from neurodegenerative conditions, in heart MR imaging [2] and freehand ultrasound [3].

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 355–362, 2010.

neighbourhood approach with a Gaussian kernel. Although this local interpolation approach is relatively computationally efficient for use during iterative refinement of slice to volume alignment, the final reconstruction can introduce additional blurring and does not make use of oversampling of the anatomy provided by the acquisition of multiple datasets. Jiang *et al.* in [5] have proposed a very similar approach for volume reconstruction based on a slice-to-volume registration and B-spline interpolation. They have applied their method on various MR brain images: awake neonates, deliberately moved adults and foetuses. To form 3D images of fetal brain, Kim *et al.* in [6] have recently proposed a new registration approach which considers the collective alignment of all slices directly, via shared structure in their intersections, rather than to an estimated 3D volume. After slice alignment a final volume is reconstructed using sparse data interpolation. All these approaches consist of two steps (motion estimation and image reconstruction) and make use of interpolation techniques to reconstruct a final 3D MR images.

In this paper we begin by assuming that fetal motion can be reliably estimated [4], and focus on the image reconstruction step and we investigate the use of a super resolution (SR) approach for this specific issue. SR in MRI has been mostly investigated using a specialized protocol to acquire shifted images, where the motion between slices is known (translation in the slice direction [7,8,9] or rotation around a common frequency encoding axis [10]). In this paper, we study reconstruction techniques with a commonly used fetal imaging protocol: multiple orthogonal stacks of thick slices where the slice thickness is usually around three times the dimension of the in-plane resolution. Since the limits of SR algorithms (maximum magnification factor, minimum number of low-resolution (LR) inputs) is still an open question [11,12], here we also experimentally investigate the number of LR images required to form a given image resolution in order to optimize the overall imaging time.

2 Problem Statement

2.1 Super-Resolution

The principle of super-resolution (SR) is to combine low resolution images to produce an image that has a higher spatial resolution than the original images [13]. This is a large research field encompassing many applications. However the majority of the work has focused on using lower resolution data acquired on a regular grid and often assuming simple translational motion between the lower resolution sample grids, unlike multislice brain MR data which is corrupted by full 3D rigid motion on a slice by slice basis (see Figure 1). Please note that this data specificity does not affect the way the SR problem is modeled. However, it introduces additional challenges because SR becomes a reconstruction problem with scattered anisotropic data.

As in most of common SR approaches, we model the physical problem and then compute a solution by inverting this observation model:

$$\mathbf{y_{r,s}} = S_r B_r W_s W_r \mathbf{x} + \mathbf{n_r} \quad for \quad 1 \leq r \leq n, \quad 1 \leq s \leq m_r \tag{1}$$

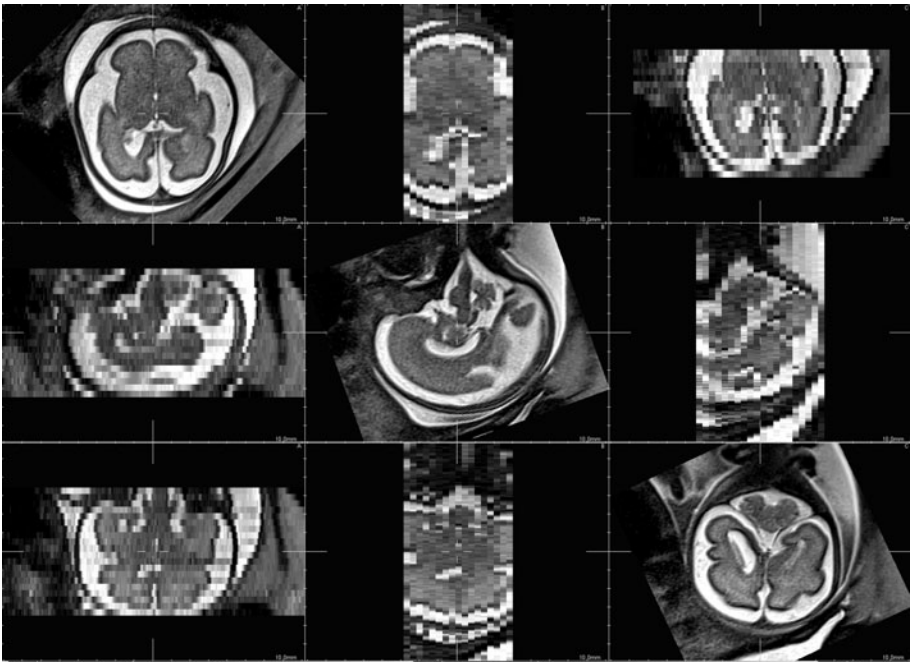


Fig. 1. Acquired fetal MR image data with anisotropic resolution. It can be noted that the LR images provide complementary views of the entire brain. From left to right: 1) axial LR image, 2) sagittal LR image, 3) coronal LR image.

where n is the number of LR images and m_r is the number of slices of the LR image r, $\mathbf{y_{r,s}}$ denotes the slice s of the LR image r, $\mathbf{x}$ is the HR image, $\mathbf{n_r}$ represents the observation noise, S_r is the subsampling matrix, B_r a blur matrix, W_s and W_r are geometric transformations of sth slice of the $\mathbf{y_{r,s}}$ and of the rth low resolution image respectively. The purpose of super-resolution is to remove the effects of the blurring convolution and to increase the voxel grid density.

For simplicity, the four operators can be combined into a single matrix $H_{r,s}$: $H_{r,s} = S_r B_r W_s W_r$. The matrix $H_{r,s}$ thus incorporates motion compensation, degradation effects, and sub-sampling for each slice $\mathbf{y_{r,s}}$. In this paper, we assume that the operators D_r and B_r, and the noise characteristics are known and spatially invariant. The operators W_s and W_r are estimated using a similar method to the one proposed by Rousseau *et al.* in [4] using the pipeline detailed in Figure 2. Using this observation model, a straightforward definition of the data discrepancy functional is:

$$D(\mathbf{x}) = \sum_{r}^{n} \sum_{s}^{m_r} \psi(\mathbf{y_{r,s}} - H_{r,s}\mathbf{x}). \tag{2}$$

ψ is usually set as a quadratic functional by assuming that the image noise $\mathbf{n_r}$ follows a Gaussian distribution. However, $\hat{\mathbf{x}}$ cannot be uniquely determined by

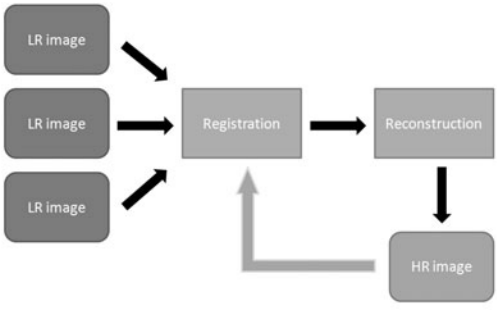


Fig. 2. Pipeline used to estimate the matrix W and the HR image $\mathbf{x}$. This pipeline is performed in an iterative fashion (this is represented by the gray arrow).

minimizing $D(\mathbf{x})$ since the image reconstruction process is an ill-posed problem. For such inverse problems, some form of regularization plays a crucial role and must be included in the cost function to stabilize the problem or constrain the space of solutions.

2.2 Variational Regularization

One of the most important classes of linear deterministic regularization methods is the one introduced by Tikhonov in [14] for which the regularization term can be defined as follows:

$$J(\mathbf{x}) = \sum_{p=0}^{P} \int c_p(\mathbf{v}) |\mathbf{x}^{(p)}(\mathbf{v})|^2 d\mathbf{v} \tag{3}$$

where weights c_p are strictly positive and $\mathbf{x}^{(p)}$ is the p^{th} order derivative of $\mathbf{x}$. Much work has been carried out on the derivation of regularization terms of the form $\phi(|\nabla \mathbf{x}|)$ or $\phi(|\mathbf{x}|)$ where ϕ is a convex even function, increasing on $\mathbb{R}^+$ (see for instance [15]). If $\phi(t) = t^2$, one obtains $\mathbf{L}^2$ norm which smooths strong variations in the image. If $\phi(t) = t$, one obtains $\mathbf{L}^1$ norm. One popular variational regularization functional relies on the total variation:

$$J_{TV}(\mathbf{x}) = \int |\nabla \mathbf{x}(\mathbf{v})| d\mathbf{v} \tag{4}$$

Charbonnier *et al.* in [15]) have proposed several edge-preserving regularization functionals whose one we use in our experiments is:

$$J_{Ch}(\mathbf{x}) = \int \phi(\nabla \mathbf{x}(\mathbf{v})) d\mathbf{v} \tag{5}$$

with $\phi(t) = 2\sqrt{1+t^2} - 2$.

3 Sufficient Number of LR Images

The principle of SR techniques is to make use of multiple LR images to estimate a single HR image. As it was pointed out in [11,12], two key points of SR algorithms are the magnification factor and the number of input images. Lin and Shum have shown in [12] that considering 2D images, the maximum magnification factor is 1.6 if the denoising and registration is not good enough and 5.7 under synthetic conditions. Moreover, when the magnification factor $\mathcal{M}$ is an integer, the sufficient number of LR images is $\mathcal{M}^2$.

In the context of fetal MR data, we consider anisotropic data where the slice thickness is approximately three times bigger than the in-plane resolution. In this work, the sufficient number of LR fetal MR images is estimated by using simulated data. To explore the ability to reconstruct a high resolution volume for typical anatomical brain structures, we applied the SR reconstruction algorithm on a T1-weighted MR image from the Brainweb dataset [16]. Brainweb is a simulated brain database which is often used as a gold standard for the analysis of in vivo acquired data. In order to mimic the clinical fetal imaging protocol used in routine, orthogonal low resolution images ($1 \times 1 \times 3$ mm) have been created using the observation model described by Equation 1. When using more than 3 images, the LR orthogonal images are simulated using 1mm shift. In this noise free case, no noise except the rounding error is introduced to the LR images and the registration is exact. We also have evaluated the performance of the proposed method using noisy low resolution images (Gaussian noise with standard deviation set to 10). Reconstruction algorithms have been applied to these synthetic images in order to reconstruct an isotropic ($1 \times 1 \times 1$ mm) image.

For quantitative comparison, the peak signal to noise ratio (PSNR) is also reported in decibels (dB): $PSNR = 10 \log_{10} \left(\frac{d^2}{|\Omega|^{-1} \sum_{\mathbf{v} \in \Omega} (\mathbf{x}(\mathbf{v}) - \hat{\mathbf{x}}(\mathbf{v}))^2} \right)$ where d is the reference image dynamic. PSNR performance with respect to the number of considered LR images are reported in Figure. Two SR regularization functionals (J_{TV}, J_{Ch}) are compared with the local scattered data interpolation approach proposed by Rousseau *et al.* in [4] (we found, in experiments not presented in this

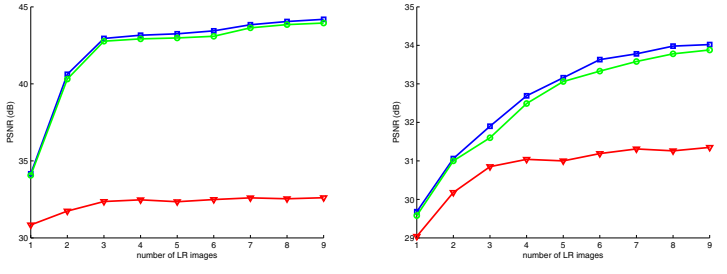


Fig. 3. PSNR in function of the number of input Brainweb images. Left: noise free experiment, Right: Gaussian noise experiment. □ : SR technique with J_{Ch} regularisation, ○: SR technique with J_{TV} regularisation, ▽: sparse interpolation technique [4].

paper, that this interpolation method has similar performance to other scattered data interpolation techniques such as those using Radial Basis Functions).

Results are reported in Figure 3. When considering the noise free experiment, it clearly appears that using more than 3 LR images does not lead to significant improvement. This means that under optimal conditions, only 3 LR images are required to accurately estimate a HR image using SR techniques. When adding non negligible noise in LR images, the number of required LR images greatly increase (except of the interpolation technique for which 3 remains a sufficient number of LR inputs) and is close to the estimated number obtained by Lin and Shum in [12]. Such experiment tends to show that efficient denoising method may have a substantial impact on SR results. In both cases, SR techniques lead to similar results and outperform the sparse interpolation approach.

4 Experiments on Fetal Brain MR Data

We applied the algorithm to fetal MR scans: T2 weighted HASTE sequence (TE/TR = 147/3190 ms) on a 1.5 T Siemens Avanto MRI Scanner (SIEMENS, Erlangen, Germany), resolution : $0.74 \times 0.74 \times 3.45$mm. Pregnant women were briefed before the exam and signed informed consent. To reduce motion artefacts, fetal sedation was obtained with 1 mg of flunitrazepam given orally to the

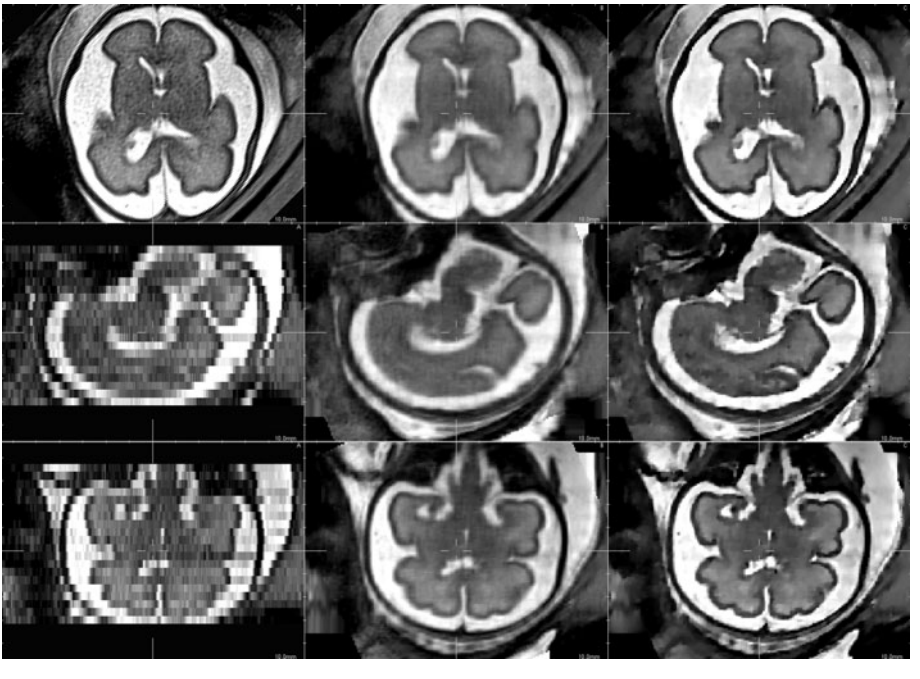


Fig. 4. Details of a reconstructed fetal brain MR image using 3 orthogonal LR images. From left to right: A) original low resolution image, B) reconstructed image using local sparse interpolation [4], C) SR reconstruction using J_{Ch} regularisation.

mother 30 minutes before the exam. The study was approved by the local ethics committee.

Figure 4 shows one original low resolution image compared to the high resolution reconstructed images for axial, coronal, sagittal views. Results obtained with the SR approach with J_{Ch} regularisation compare favorably with the local sparse interpolation approach proposed by Rousseau *et al.* in [4]. It can be especially noticed that the boundaries of brain structures (like the cortex) are better recovered. Moreover, the noise present in the LR image has no major impact on the reconstructed images. However, reconstruction artefacts appear outside the fetal head (see for instance the surrounding structures). In this case, the assumption concerning the "rigidity" of the scene is violated and registration errors disturb the reconstruction process. The result shown in Figure 4 has been obtained using 3 LR images. Using up to 6 images gives only marginal qualitative improvements. This experiment tends to show that the pipeline described in Figure 2 may be efficient enough to reduce, in our context, fetal imaging time by avoiding unnecessary acquisitions.

5 Discussion

The first contribution of this work concerns the use of a SR framework for HR 3D isotropic fetal brain MR image reconstruction. Contrary to previous works [4,5,6] where sparse interpolation frameworks were used, an observation model is introduced in this paper in order to truly take into account physics of MR image acquisition (point spread function, sub-sampling and noise). This model allows potentially to recover fine image details compared to interpolation based approaches. Visual analysis of the obtained results on real data are very encouraging. Such high resolution image reconstruction algorithm represents an important step towards analysis of fine scale anatomical details.

The second contribution of this paper is the study of the sufficient number of LR images based on simulations using the Brainweb images. Considering a set of orthogonal stack of 2D slices with an anisotropic resolution of $1 \times 1 \times 3$mm, the sufficient number of LR images is three in the case of noise free images. When adding Gaussian noise, the sufficient number of LR inputs is less than 9. Including an efficient denoising step in the overall image processing pipeline should lead to a significant decrease of the number of required LR images. As a consequence, further work is to investigate the building of such pipeline in order to reduce the fetal imaging time.

Acknowledgment. The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013 Grant Agreement no. 207667). This work is also funded by NIH/NINDS Grant R01 NS 055064 and a CNRS grant for collaboration between LSIIT and BICG.

References

1. Prayer, D., Brugger, P.C., Prayer, L.: Fetal MRI: techniques and protocols. Pediatric Radiology 34(9), 685–693 (2004)
2. Moore, J., Drangova, M., Wierzbicki, M., Barron, J., Peters, T.: A high resolution dynamic heart model based on averaged MRI data. In: Ellis, R.E., Peters, T.M. (eds.) MICCAI 2003. LNCS, vol. 2878, pp. 549–555. Springer, Heidelberg (2003)
3. Prager, R.W., Gee, A., Berman, L.: Stradx: real-time acquisition and visualization of freehand three-dimensional ultrasound. Medical Image Analysis 3(2), 129–140 (1999)
4. Rousseau, F., Glenn, O.A., Iordanova, B., Rodriguez-Carranza, C., Vigneron, D.B., Barkovich, J.A., Studholme, C.: Registration-based approach for reconstruction of high-resolution in utero fetal MR brain images. Academic Radiology 13(9), 1072–1081 (2006)
5. Jiang, S., Xue, H., Glover, A., Rutherford, M., Rueckert, D., Hajnal, J.V.: MRI of moving subjects using multislice snapshot images with volume reconstruction (SVR): application to fetal, neonatal, and adult brain studies. IEEE Transactions on Medical Imaging 26(7), 967–980 (2007)
6. Kim, K., Habas, P.A., Rousseau, F., Glenn, O.A., Barkovich, A.J., Studholme, C.: Intersection based motion correction of multislice MRI for 3-D in utero fetal brain image formation. IEEE Transactions on Medical Imaging 29(1), 146–158 (2010)
7. Carmi, E., Liu, S., Alon, N., Fiat, A., Fiat, D.: Resolution enhancement in MRI. Magnetic Resonance Imaging 24(2), 133–154 (2006)
8. Peeters, R.R., Kornprobst, P., Nikolova, M., Sunaert, S., Vieville, T., Malandain, G., Deriche, R., Faugeras, O., Ng, M., Hecke, P.V.: The use of super-resolution techniques to reduce slice thickness in functional MRI. International Journal of Imaging Systems and Technology 14(3), 131–138 (2004)
9. Greenspan, H., Peled, S., Oz, G., Kiryati, N.: MRI inter-slice reconstruction using Super-Resolution. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 1204–1206. Springer, Heidelberg (2001)
10. Shilling, R.Z., Robbie, T.Q., Bailloeul, T., Mewes, K., Mersereau, R.M., Brummer, M.E.: A super-resolution framework for 3-D high-resolution and high-contrast imaging using 2-D multislice MRI. IEEE Transactions on Medical Imaging 28(5), 633–644 (2009)
11. Baker, S., Kanade, T.: Limits on Super-Resolution and how to break them. IEEE Trans. Pattern Anal. Mach. Intell. 24(9), 1167–1183 (2002)
12. Lin, Z., Shum, H.: Fundamental limits of Reconstruction-Based superresolution algorithms under local translation. IEEE Trans. Pattern Anal. Mach. Intell. 26(1), 83–97 (2004)
13. Bose, N., Chan, R., Ng, M.: Special issue: High resolution image reconstruction. International Journal of Imaging Systems and Technology 14, 2–3 (2004)
14. Tikhonov, A.: Regularization of incorrectly posed problems. Sov. Math. Dokl. 4, 1624–1627 (1963)
15. Charbonnier, P., Blanc-Feraud, L., Aubert, G., Barlaud, M.: Deterministic edge-preserving regularization in computed imaging. IEEE Transactions on Image Processing 6(2), 298–311 (1997)
16. Cocosco, C.A., Kollokian, V., Kwan, R.K., Evans, A.C.: BrainWeb: online interface to a 3D MRI simulated brain database (1997)

Spatial Regularization of Functional Connectivity Using High-Dimensional Markov Random Fields

Wei Liu[1], Peihong Zhu[1], Jeffrey S. Anderson[2], Deborah Yurgelun-Todd[3], and P. Thomas Fletcher[1]

[1] Scientific Computing and Imaging Institute, University of Utah, USA
weiliu@sci.utah.edu
[2] Department of Radiology, University of Utah, USA
[3] Department of Psychiatry, University of Utah, USA

Abstract. In this paper we present a new method for spatial regularization of functional connectivity maps based on Markov Random Field (MRF) priors. The high level of noise in fMRI leads to errors in functional connectivity detection algorithms. A common approach to mitigate the effects of noise is to apply spatial Gaussian smoothing, which can lead to blurring of regions beyond their actual boundaries and the loss of small connectivity regions. Recent work has suggested MRFs as an alternative spatial regularization in detection of fMRI activation in task-based paradigms. We propose to apply MRF priors to the computation of functional connectivity in resting-state fMRI. Our Markov priors are in the space of pairwise voxel connections, rather than in the original image space, resulting in a MRF whose dimension is twice that of the original image. The high dimensionality of the MRF estimation problem leads to computational challenges. We present an efficient, highly parallelized algorithm on the Graphics Processing Unit (GPU). We validate our approach on a synthetically generated example as well as real data from a resting state fMRI study.

1 Introduction

Functional magnetic resonance imaging (fMRI) provides a non-invasive measurement of cerebral blood flow in the brain that can be used to infer regions of neural activity. Traditional fMRI studies are based on measuring the response to a set of stimuli, and analysis involves testing the time series at each voxel for correlations with the experimental protocol. Recently, there has been growing interest in using resting-state fMRI to infer the connectivity between spatially distant regions. A standard approach is to use correlation between pairs of time series as a measurement of their functional connectivity. The high level of noise present in fMRI can cause errors in pairwise connectivity measurements, resulting in spurious false connections as well as false negatives.

In both task-based and resting-state fMRI the impact of imaging noise can be reduced by taking advantage of the spatial correlations between neighoring voxels

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 363–370, 2010.

in the image. A common approach used for instance in Statistical Parametric Mapping (SPM)[1] is to apply a spatial Gaussian filter to smooth the signal prior to statistical analysis. However, this can lead to overly blurred results, where effects with small spatial extent can be lost and detected regions may extend beyond their actual boundaries. An alternative approach to spatial regularization that has been proposed for task activation paradigms is to use a Markov Random Field (MRF) prior [2,3,4,5,6], which models the conditional dependence of the signals in neighboring voxels.

In this work we propose to use MRF models in resting-state fMRI to leverage spatial correlations in functional connectivity maps. Unlike previous MRF-based approaches, which use the neighborhood structure defined by the original image voxel grid, the neighborhoods in functional connectivity must take into account the possible relationships between spatially distant voxels. Therefore, we define the neighborhood graph on the set of all voxel pairs. This results in a Markov structure on a grid with twice the dimensions of the original image data, i.e., the pairwise connectivities for three-dimensional images results in a six-dimensional MRF. The neighborhood structure is defined so that two voxels are more likely to be connected if they are connected to each other's spatial neighbors.

We combine the Markov prior on functional connectivity maps with a likelihood model of the time series correlations in a posterior estimation problem. Furthermore, we model solve for the unknown parameters of the MRF and likelihood using an Expectation Maximization (EM) algorithm. In the estimation step the posterior random field is sampled using Gibbs Sampling and estimated using Mean Field theory.

In the next section we describe our MRF model of functional connectivity maps. In Section 3 we give the details of the algorithm to estimate the functional connectivity probabilities, including implementation details for the GPU solver. Finally, in Section 4 we demonstrate the advantages of our approach on a synthetically generated data set as well as on real resting-state fMRI data.

2 Markov Random Fields for Functional Connectivity

Our framework for functional connectivity is a Bayesian approach in which we estimate the posterior distribution of the connectivity between voxels, conditioned on the fMRI data. Let $\boldsymbol{x} = \{x_{ij}\}$ denote the functional connectivity map, i.e., a map denoting whether each pair of voxels i, j is connected, and let $\boldsymbol{y}$ denote the original fMRI data, or some meausrements derived from the fMRI. In this work we take $\boldsymbol{y}$ to be the map of correlations between pairs voxel time series. The posterior distribution is then proportionally given by

$$P(\boldsymbol{x} \,|\, \boldsymbol{y}) \propto P(\boldsymbol{x}) \cdot P(\boldsymbol{y} \,|\, \boldsymbol{x}). \tag{1}$$

In this work we model $P(\boldsymbol{x})$, the prior on the connectivity map, using a MRF, and the likelihood $P(\boldsymbol{y} \,|\, \boldsymbol{x})$ using Gaussian models of the Fisher transformed correlation values. We now give details for both of these pieces of the model.

2.1 Markov Prior

Conventional image analysis applications of MRFs [7] define the set of sites of the random field as the image voxels, with the neighborhood structure given by a regular lattice. Because we are studying the pairwise connectivity between voxels, we need to define a MRF in the higher-dimensional space of voxel location pairs. Thus, if $\Omega \subset \mathbb{Z}^d$ is a d-dimensional image domain, then the sites for our connectivity MRF form the set $\mathcal{S} = \Omega \times \Omega$. Let $i, j \in \Omega$ be voxel locations, and let $\mathcal{N}_i, \mathcal{N}_j$ denote the set of neighbors of voxel i and j, respectively, in the original image lattice. Then the set of neighbors for the site $(i, j) \in \mathcal{S}$ is given by $\mathcal{N}_{ij} = (\{i\} \times \mathcal{N}_j) \cup (\mathcal{N}_i \times \{j\})$. In other words, two sites are neighbors if they share one coordinate and their other coordinates are neighbors in the original image lattice. This neighborhood structure will give us the property in the MRF that two voxels i, j in the image are more likely to be connected if i is connected to j's neighbors or vice-versa. Equipped with this graph structure, $\mathcal{S}$ is a regular $2d$-dimensional lattice, which we will refer to as the *connectivity graph*.

We next define a multivariate random variable $\boldsymbol{x} = \{x_{ij}\}$ on the set $\mathcal{S}$, where each random variable x_{ij} is a binary variable that denotes the connectivity ($x_{ij} = 1$) or lack of connectivity ($x_{ij} = -1$) between voxel i and j. If $A \subset \mathcal{S}$, let $\boldsymbol{x}_A$ denote the set of all x_{ij} with $(i, j) \in A$, and let $\boldsymbol{x}_{-ij}$ denote the collection of all variables in $\boldsymbol{x}$ excluding x_{ij}. For $\boldsymbol{x}$ to be a MRF it must satisfy

$$P(x_{ij} \,|\, \boldsymbol{x}_{-ij}) = p(x_{ij} \,|\, x_{\mathcal{N}_{ij}}).$$

According to the Hammersley and Clifford Theorem[8], $\boldsymbol{x}$ is Markov random field if and only if it is also a Gibbs distribution, defined as

$$P(\boldsymbol{x}) = \frac{1}{Z} \exp\left(-U(\boldsymbol{x})\right), \tag{2}$$

where U is the energy function $U(\boldsymbol{x}) = \sum_{c \in \mathcal{C}} V_c$, with potentials V_c defined for each clique c in the clique set $\mathcal{C}$. The partition function $Z = \sum \exp(-U(\boldsymbol{x}))$ is a normalizing constant, where the summation is over all possible configurations of $\boldsymbol{x}$. We use a particular form of MRF—the Ising model—a commonly used model for MRFs with binary states. In this model the energy function is given by

$$U(\boldsymbol{x}) = -\beta \sum_{\langle ij,mn \rangle} x_{ij} x_{mn}, \tag{3}$$

where the summation is over all edges $\langle ij, mn \rangle$, i.e., all adjacent voxel pairs $(i, j), (m, n)$ in the connectivity graph. When $\beta > 0$, this definition favors similarity of neighbors.

2.2 Likelihood Model

We now define the likelihood model, $P(\boldsymbol{y} \,|\, \boldsymbol{x})$, which connects the observed data $\boldsymbol{y}$ to our MRF. Because we are interested in the functional connectivity between pairs of voxels, we compute the correlation between the time courses of each

pair of voxels, and get a correlation matrix $\boldsymbol{y} = \{y_{ij}\}$. Just as in the case of the random field $\boldsymbol{x}$, the correlation matrix $\boldsymbol{y}$ is also defined on the 2-dimensional lattice $\mathcal{S}$. Linear correlation is not the only choice for $\boldsymbol{y}$. We can use any statistic, as long as it indicates the affinity between two voxel time series. Another possibility could be frequency domain measures, such as the coherence [9].

Before defining the full data likelihood, we start with a definition of the *emission function* at a single site $s_{ij} \in \mathcal{S}$. This is defined as the conditional likelihood, $P(y_{ij} \,|\, x_{ij})$, and is interpreted as the probability of the observed correlation, given the state of the connectivity between voxels i and j. We model the emission function as a Gaussian distribution with unknown mean and variance on the Fisher transformed correlation y_{ij}, that is,

$$P(y_{ij} \,|\, x_{ij} = k) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(F(y_{ij}) - \mu_k)^2}{2\sigma_k^2}\right), \tag{4}$$

where F denotes the Fisher transform. Notice that each correlation y_{ij} on site s_{ij} only depends on the latent variable x_{ij} on the same site, and does not depend on neighbors of x_{ij}. Therefore, the full likelihood is given by

$$P(\boldsymbol{y} \,|\, \boldsymbol{x}) = \prod_{s_{ij} \in \mathcal{S}} P(y_{ij} \,|\, x_{ij}). \tag{5}$$

3 Estimation via Expectation Maximization

Having defined the data likelihood and MRF prior in the previous section, we are now ready to describe the maximization of the posterior given by (1). For this we need to determine the model parameters, β in (3) and (μ_k, σ_k^2) in (4). Rather than arbitrarily setting these parameters, we estimate them using an Expectation Maximization (EM) algorithm. Exact computation of the full posterior (1) is intractable, due to the combinatorial number of terms in the partition function Z. Therefore, we instead maximize the approximate posterior given by the pseudo-likelihood function [7,8],

$$PL(\boldsymbol{x}, \boldsymbol{y}) = \prod_{ij} P(x_{ij} | x_{\mathcal{N}_{ij}}) P(y_{ij} | x_{ij}). \tag{6}$$

In the E-step, the parameters are held fixed and we compute the posterior probability for each x_{ij}, and sample x_{ij} from the posterior distribution using Gibbs Sampling. We then compute the expected value of each connectivity node by Mean Field theory. After we compute the posterior of current point x_{ij}, we update x_{ij} with its expected value $\langle x_{ij} \rangle$.

In the M-step, the complete data $\{\langle \boldsymbol{x} \rangle, \boldsymbol{y}\}$ is available, and the parameters can be estimated by maximizing the joint pseudo-likelihood given by (6) using Newton's method. After several iterations of this EM algorithm, we get parameters as our MAP estimates.

GPU Implementation. The whole algorithm involves updating a high dimensional connectivity matrix $\boldsymbol{x}$ iteratively, and hence have high computation cost. We designed a parallel Markov random field updating strategy on graphics processing unit (GPU). The algorithm take only a few minutes compared with more than 1 hour on CPU counterpart.

To fit the algorithm into GPU's architecture, we use some special strategy. First, because GPU only support 3-dimensional array, we need to reshape $\boldsymbol{x}$ and $\boldsymbol{y}$ defined originally on higher dimensional graph by linear indexing their original subscripts. This is especially difficult for brain fMRI data because the gray matter voxels resides in a irregular 3-D lattice. Specific data structure are used for mapping between original voxel subscripts and their linear index i and j. Second, to update each site of MRF in parallel we have to make sure a site is not updated simultaneously with its neighbors, otherwise the field tends to be stuck in a checkerboard-like local maximum. Our strategy is to divide all the sites of the field into several sub-groups, such that a site is not in the same sub-group with its neighbors. We then can update the sub-group sequentially, while the data in sub-groups are updated simultaneously. The whole procedure is summarized in Algorithm 1.

Algorithm 1. MAP estimation by EM

Require: Sample correlation matrix $\mathbf{Y}$.
Init posterior matrix by maximizing conditional likelihood $P(y_{ij}|x_{ij})$.
while $\Delta\{\beta, \mu, \sigma^2\} > \varepsilon$ **do**
E step:
(a) Based on the current parameters, compute the posterior by (6).
(b) Repeatedly Do Gibbs Sampling until the field stabilize.
(c) Based on current value of x_{ij}, iteratively compute the mean field for all nodes in $\mathcal{S}$ until the field stable.
M step:
(d) With complete data $\{\mathbf{X}, \mathbf{Y}\}$, estimate β , μ and σ^2 by maximizing (6).
end while
return posterior matrix $\mathbf{X}$.

4 Results

Synthetic Data. We first construct a synthetic data set consisting of a 100×1 1-D image, with each pixel a 300-point time course signal. The time course was constructed with a baseline DC signal of 800, plus additive Gaussian noise of variance 50. We then added a sine wave signal of frequency 0.2 and amplitude 20 to two distant regions of the image. The goal is to detect the connectivity between these two distant regions. Between those pixels containing signal the connectivity is 1, otherwise it is 0 for connectivity between signal and noise, and between noise time series. The true connectivity map is shown in Fig. 1(a)

To compare our MRF model with conventional Gaussian blurring of the correlation map, we applied both approaches to the synthetic data (Fig. 1). On the

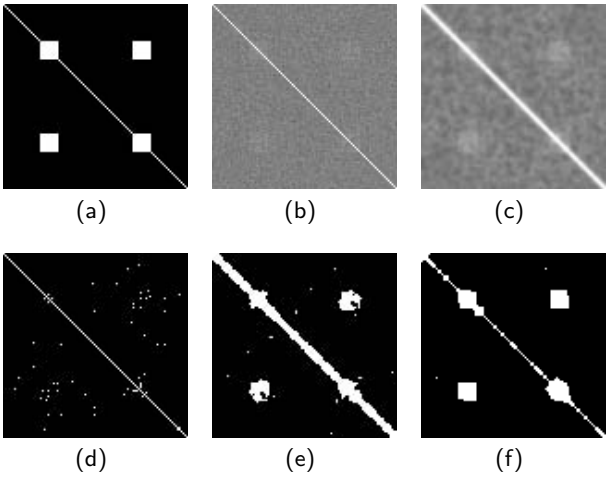


Fig. 1. Test of synthetic data, showing the (a) ground-truth connectivity, (b) correlation of original, noisy data, (c) correlation of Gaussian-smoothed data, (d) connectivity based on noisy correlations, (e) connectivity based on smoothed data, (f) connectivity computed using proposed MRF model

correlation map in the top row, we see smoothing does remove noise and results in a correlation map that looks more like the true connectivity map. However, it also creates several errors, most noticeably false positives around the diagonal (Fig. 1(e)). Fig. 1(f) shows the proposed MRF method better detects the true connectivity regions while removing most false positives.

Resting-State fMRI. Next we tested our method on real data from healthy control subjects in a resting-state fMRI study. BOLD EPI images (TR= 2.0 s, TE = 28 ms, GRAPPA acceleration factor = 2, 40 slices at 3 mm slice thickness, 64 x 64 matrix, 240 volumes) were acquired on a Siemens 3 Tesla Trio scanner with 12-channel head coil during the resting state, eyes open. The data was motion corrected by SPM software and registered to a T2 structural image. We used a gray matter mask from an SPM tissue segmentation so that only gray matter voxels are counted in the connectivity analysis. We do *not* spatially smooth the data, in order to see the benefit of replacing spatial smoothing with our MRF method. Before computing the correlations, the time series at all voxels are linearly detrended by least squares regression.

Fig. 2 compares the real data results using no spatial regularization, Gaussian smoothing, and the proposed MRF model. Though the posterior connectivity of the MRF is computed between every pair of voxels within a slice, for visualization purposes, only the posterior of the connectivity between one voxel and the slice is shown. We chose to visualize the connectivity to a voxel in the posterior cingulate cortex (PCC) because this is known to be involved in the Default Mode Network [10], with connections to the medial prefrontal cortex (MPFC). The results show that Gaussian smoothing is able to remove noise, but is unable

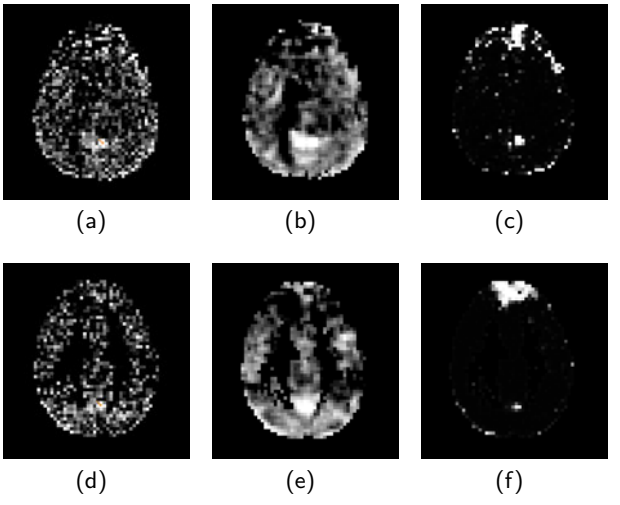


Fig. 2. Correlation map and Posterior Connectivity map between seed voxel and slice containing the seed. First row is subject 1. (a) the correlation map computed from data without spatial smoothing. (b) correlation map of data after smoothing. (c) Posterior probability computed from MRF. Second row (d,e,f) is subject 2 with same test.

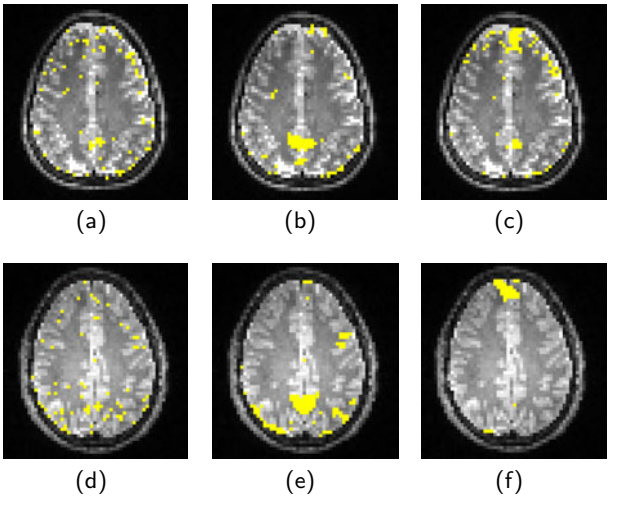


Fig. 3. Thresholded correlation map and Posterior Connectivity map between seed voxel and slice, overlaid to T2 image. First row is subject 1. (a) the correlation map computed from data without spatial smoothing. (b) After smoothing. (c) Posterior probability by MRF. Second row (d,e,f) is subject 2 with same test.

to find a clear connection between the PCC and the MPFC. Our proposed MRF model (rightmost plot) is able to remove spurious connections, and also clearly shows a connection to the MPFC.

To show the strongest connections found by each method, Fig. 3 shows the thresholded connectivity maps overlaid on T2 structural image. Images in the first two columns are thresholded such that the top 5% voxel correlations are shown. For the MRF in the third column, the MAP estimate is shown.

5 Conclusion

We propose a Markov random field and Bayesian framework for spatially regularizing functional connectivity. Future work may include running this pairwise connectivity analysis on 3D whole brain. Another interesting direction is applying the method to multiple sessions and subjects.

Acknowledgments

This work was funded in part by NIH Grant R01 DA020269 (Yurgelun-Todd).

References

1. Worsley, K.J., Friston, K.J.: Analysis of fMRI time-series revisited–again. Neuroimage 2(3), 173–181 (1995)
2. Ou, W., Golland, P.: From spatial regularization to anatomical priors in fMRI analysis. Information in Medical Imaging, 88–100 (2005)
3. Descombes, X., Kruggel, F., Cramon, D.V.: Spatio-temporal fMRI analysis using Markov random fields. IEEE Trans. Medical Imaging 17(6), 1028–1039 (1998)
4. Descombes, X., Kruggel, F., von Cramon, D.Y.: fMRI signal restoration using a spatio-temporal Markov random field preserving transitions. NeuroImage 8(4), 340–349 (1998)
5. Woolrich, M.W., Jenkinson, M., Brady, J.M., Smith, S.M.: Fully Bayesian spatio-temporal modeling of fMRI data. IEEE Transactions on Medical Imaging 23(2), 213–231 (2004)
6. Cosman, E.R., Fisher, J.W., Wells, W.M.: Exact MAP activity detection in fMRI using a GLM with an ising spatial prior. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 703–710. Springer, Heidelberg (2004)
7. Li, S.Z.: Markov Random Field Modeling in Image Analysis. Springer, Heidelberg (2009)
8. Besag, J.: Spatial interaction and the statistical analysis of lattice systems. Journal of the Royal Statistical Society. Series B (Methodological) 36(2), 192–236 (1974)
9. Müller, K., Lohmann, G., Bosch, V., von Cramon, D.Y.: On multivariate spectral analysis of fMRI time series. NeuroImage 14(2), 347–356 (2001)
10. Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L.: A default mode of brain function. PNAS 98(2), 676–682

Shell Model for Reconstruction and Real-Time Simulation of Thin Anatomical Structures

Olivier Comas[1,2], Christian Duriez[1], and Stéphane Cotin[1]

[1] INRIA, Shaman team, Lille, France
[2] Preventative Health Flagship, CSIRO ICT, AEHRC, Brisbane, Australia

Abstract. This paper presents a new modelling technique for the deformation of thin anatomical structures like membranes and hollow organs. We show that the behaviour of this type of surface tissue can be abstracted with a modelling of their elastic resistance using shell theory. In order to apply the shell theory in the context of medical simulation, our method propose to base the geometrical reconstruction of the organ on the shape functions of the shell element. Moreover, we also use these continuous shape functions to handle the contacts and the interactions with other types of deformable tissues. The technique is illustrated using several examples including the simulation of an angioplasty procedure.

1 Introduction

The human body is composed of various deformable anatomical structures. A key challenge of soft-tissue modelling is the variousness of the mechanical behaviours. It seems unrealistic to use a unique model for all tissues. Yet most of previous works focus on volumetric models that are able to capture the behaviour of solid organs like the liver or the brain (see for instance [1,2]). In contrast, this paper seeks to propose a solution for simulating, in real-time, the deformation of thin anatomical structures whose volume is negligible compared to their surface area. Examples include hollow structures, such as the wall of blood vessels, or membranes, such as the Glisson's capsule surrounding the liver. It is also of particular interest to us for modelling the colon in our colonoscopy simulator [3].

Shell theory allows the modelling of structure deformations when the thickness is small compared to its other dimensions [4]. The key idea is to model the physical shell as a surface but endowed with mechanical properties in the form of elastic resistance to stretching and bending forces. Rather than resorting to shell theory, previous works in medical simulation often rely on linear or angular mass-spring models as in [5,6]. Yet, such models are limited in their ability to describe certain behaviour, as they do not rely on continuum mechanics: it is difficult to derive spring stiffness (in particular for angular springs) from elastic properties (Young's modulus and Poisson's ratio). The work of Choi et al. [7] in computer graphics refers to a fast shell-based modelling for interactive simulation. Their model relies on simplified energy functions and precomputed modal analysis for fast and visually realistic results. We propose to rely on a similar approach but

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 371–379, 2010.

with more accuracy to be applicable to medical simulation. Our model is not based on modal analysis but uses a co-rotational formulation and polynomial shape functions presented in [8].

To model the deformation of complex anatomical structures using shell elements, the first step is to describe its surface with curved patches. This process is quite similar to the reconstruction of the surface of objects in computer vision. Indeed calculating curvature maps of 3D surfaces represented as digitised data (point clouds or triangulated meshes) has been extensively studied. One of the most common approach is to use continuous surface patches [9]. The surface is locally approximated by an analytic representation called surface patch, usually chosen to be an implicit surface. These works target approaches that are not noise-sensitive and coherent surface and curve extraction from 3D data [10]. However, our situation is substantially different as we want to model the deformation of the structure. In that regard the curvature of the surface has a physical meaning: it represents the mid-surface of the shell. We propose to approximate the surface of anatomical structures with shell elements whose each surface is described by the shape function used in our shell formulation.

These polynomial shape functions are used in three different ways in our computational model: (a) to approximate complex geometrical shapes, (b) to compute internal forces, (c) to compute contact forces onto a curved triangle. Section 2 presents our Finite Element Modelling (FEM) for shell elements and how we process contacts and interactions with other models. In section 3 we introduce an automatic process to obtain meshes from image based reconstruction. Finally the benefits of our approach (meshing of a curved surface, fast computation and possible interactions with solid models) are illustrated using various examples showed in section 4. Implementations were carried out within the open source framework SOFA [11].

2 Co-rotational Triangular Shell Model for Thin Structures

A complete description and validation of our co-rotational triangular shell finite element model is available in one of our previous publication [8]. Therefore we will only remind the key points. We improved and extended a plate model first introduced by Przemieniecki [12] to a co-rotational formulation. Co-rotational approaches offer a good trade-off between computational efficiency and accuracy by allowing small deformations but large displacements. Once combined with an in-plane membrane formulation we obtain an accurate, yet computationally efficient, shell finite element method featuring both membrane and bending energies. In the following we detail the bending stiffness computation in order to present the polynomial shape functions that are used in the shell model.

Polynomial shape function. To calculate the stiffness matrix for the transverse deflections and rotations shown on Fig. 1, the deflection u_z is computed using a polynomial interpolation:

$$u_z = c_1 + c_2x + c_3y + c_4x^2 + c_5xy + c_6y^2 + c_7x^3 + c_8xy^2 + c_9y^3 \qquad (1)$$

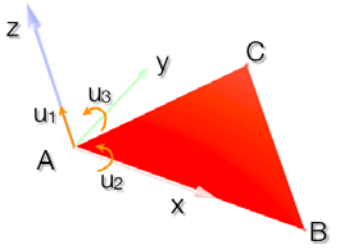


Fig. 1. The different degrees of freedom u of a triangular thin plate in bending

where $c_1, \ldots, c_9$ are constants. Using a third-degree polynomial expression allows us to reach a greater precision for both the computation of the bending energy and the interpolation within the surface of the shell. Let us define the vector $\mathbf{u} = \{u_1 u_2 \ldots u_9\}$ of the displacements and slopes at the three corners of the triangular plate using the following notations:

$$u_1 = (u_z)_{x_1,y_1} \qquad u_2 = \left(\frac{\partial u_z}{\partial y}\right)_{x_1,y_1} \qquad u_3 = -\left(\frac{\partial u_z}{\partial x}\right)_{x_1,y_1} \tag{2}$$

and so on for the two other vertices and we can derive a matrix $\mathbf{C}$ such as $\mathbf{u} = \mathbf{Cc}$ where $\mathbf{c} = \{c_1 c_2 \ldots c_9\}$. We can then calculate the strains from the flat-plate theory using:

$$e_{xx} = -z\frac{\partial^2 u_z}{\partial x^2} \qquad e_{yy} = -z\frac{\partial^2 u_z}{\partial y^2} \qquad e_{xy} = -2z\frac{\partial^2 u_z}{\partial x \partial y} \tag{3}$$

Symbolically this may be expressed as $\mathbf{e} = \mathbf{Dc}$ where $\mathbf{D}$ derives from (1) and (3). Noting that $\mathbf{c} = \mathbf{C}^{-1}\mathbf{u}$, we have $\mathbf{e} = \mathbf{DC}^{-1}\mathbf{u} = \mathbf{bu}$ where the strain-displacement matrix $\mathbf{b} = \mathbf{DC}^{-1}$. The stiffness matrix $\mathbf{K}_e$ for an element is then obtained from:

$$\mathbf{K}_e = \int_v \mathbf{b}^T \boldsymbol{\chi} \mathbf{b} dV \quad \text{where } \boldsymbol{\chi} \text{ is the material matrix .} \tag{4}$$

Mechanical interactions with the curved surface of shells. The practical interest of modelling complex behaviours such as bending and twisting would remain fairly low for medical simulation if contacts and constraints were not handled properly. In our case the difficulty comes from different sources. First the collision detection must be carried out with the curved surface of shell elements as opposed to the classic detection on plane triangles. Then forces applied to a given triangle need to be distributed between linear forces and torques onto its three vertices. As we will see, the same polynomial interpolation function chosen to compute the bending energy in our FEM formulation is also used to capture the interactions between the curved surface and other objects.

In order to detect the collision with the bent surface, we have chosen the subdivision approach. We first sample the flat surface of each element by recursively dividing each triangle into four smaller ones and the deflection of each new vertex is computed using (1) according to the displacements and slopes at the three vertices of the triangular element. This process of subdivision allows us to render each shell as a curved triangle (Fig. 2 (a) and (b)) and detect any collision with the curved surface of the shell using any of the classic collision detection algorithms working on flat triangles.

Once a collision has been detected, it must be processed by distributing the linear force received on the bent surface between the three vertices of the triangle. First the linear part of the force is simply transmitted on each node using the barycentric coordinates of the contact point's projection onto the triangle.

The main difficulty is to convert the normal component of the force applied to the bent surface into a torque at each of the three nodes (Fig. 2 (c)). Our approach is the following: during force computation, we use the change in orientation measured at each node to compute the local deflection of each subvertex within the triangle. Differentiating the formulation twice yields a relation between the torque applied at each node and the generated force in bending. We therefore need to invert the latter formulation to convert a bending force into torques at each vertex. We start by retrieving the normal component of the applied force vector F_z. We project the application point of the force into the triangle's plane and compute its local coordinates (x, y). We create the polynom $P = F_z(1 \;\; x \;\; y \;\; x^2 \;\; xy \;\; y^2 \;\; x^3 \;\; xy^2 \;\; y^3)^T$. The moments at each vertex are then obtained with $\Omega = (\mathbf{C}^{-1})^T P$. Thus we are able to transmit any force coming from interactions with the curved surface of shells to the mechanical vertices used in our FEM formulation.

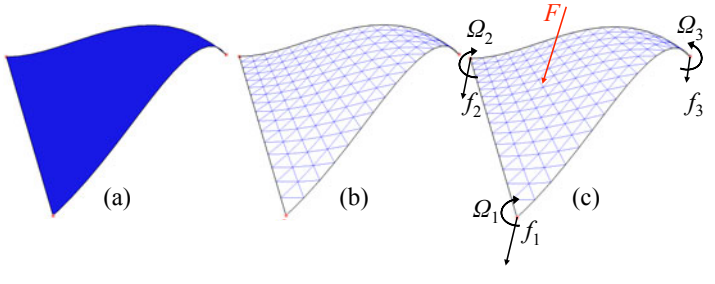


Fig. 2. (a,b) The triangle formed by the three vertices of the shell has been recursively subdivided 3 times to allow more accurate rendering and collision detection. (c) The shape function is used to distribute an external force F onto the triangle nodes.

3 Physics-Based Reconstruction Using Shell Elements

Because the surface of an anatomical structure has a physical meaning, we propose to patch the surface with triangular elements whose interpolation makes

use of the same shape function designed for our shell FEM formulation. Moreover, while many flat triangles are required to describe highly curved surfaces, fewer triangular shell elements are needed to describe the given geometry with the same precision since they can be curved. In the following we assume that we have a high resolution triangular mesh obtained from a binary segmented image of the organ we want to simulate (via a Marching Cube algorithm for instance). Our goal is to create a mesh featuring the optimal number of shell elements while staying as close as possible to our targeted geometry.

Therefore we need to ensure that the distance between the surface of our shell-based mesh and the targeted high resolution mesh will be minimal. An efficient technique for measuring the error between two surfaces is the Hausdorff distance [13]. As a reminder the Hausdorff distance between two meshes is the maximum between the two so-called one-sided Hausdorff distances:

$$d_{\mathrm{H}}(X,Y) = \max\left\{ \sup_{x\in X} \inf_{y\in Y} d(x,y), \sup_{y\in Y} \inf_{x\in X} d(x,y) \right\} \quad . \tag{5}$$

where $d()$ is the Euclidian distance between two points. The same technique of subdivision used for rendering allows us to sample the actual surface described by the shells to compute the Hausdorff distance with the targeted high resolution mesh.

The first step in the process of generating a shell-based mesh is an important decimation of the high resolution mesh, using quadric edge collapse technique implemented in Meshlab [14]. The algorithm tries as much as possible to preserve mesh boundaries and generates high quality triangular elements. We then apply a heuristic method derived from the work of Saupin et. al [15] with tetrahedral meshes based on simple geometrical rules. For each node of the coarse mesh, we find the three closest triangles on the high resolution mesh and we move the node to the barycenter of the three centres of mass of those triangles. This technique locally smoothes the surface of the mesh while converging towards the desired high resolution mesh. At each iteration of this algorithm we measure the error between the curved surface of shells and the target using the Hausdorff distance

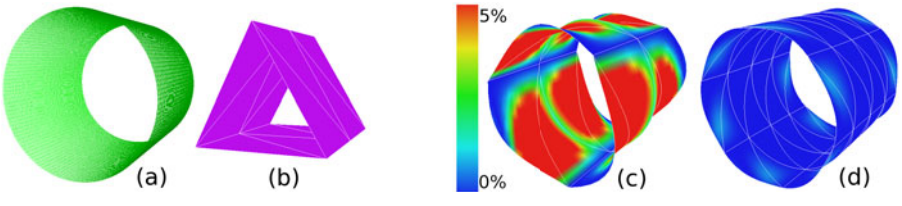


Fig. 3. The target (a) is a high resolution cylinder mesh of 16,384 triangles and we start from a very coarse mesh (12 triangles), rendered with flat triangles here (b). In (c) the coarse mesh is rendered with shells and a one-sided Hausdorff distance colour map is applied to show the initial error with the high resolution mesh. (d) One-sided Hausdorff distance colour map after one iteration of our algorithm (48 shells).

and the process is stopped when the required precision has been reached. A simple example is shown Fig. 3 to illustrate the method.

4 Results

Meshing of anatomical structures. This approach has been applied to approximate more complex anatomical geometries with curved shell elements. In each case the error is expressed as a percentage of the diagonal of the object's bounding box.

Computation times. We perform several tests on the aneurysm model at different resolutions to measure computation times (Fig. 6). The shells are resisting to a uniform pressure load and solved using a Conjugate Gradient (CG) iterative solver. Implicit integration allows for large time steps (40ms) and the computation is real-time for 800 shell elements and a reasonable error criterion (5%). When the computation time must be bounded (critical real-time applications), one can fix the number of CG iterations to, for instance, 100 and remains real-time for 1000 shell elements. However, in that case the accuracy of the results is not checked.

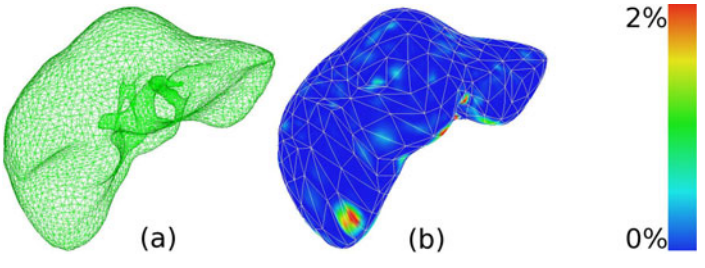


Fig. 4. (a) the targeteted high resolution Glisson's capsule mesh (8,000 triangles). (b) the one-sided Hausdorff distance error map after applying only one iteration of our algorithm to the coarse mesh (1,200 shells).

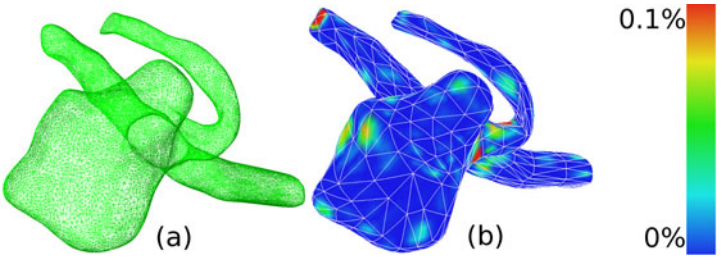


Fig. 5. (a) the targeteted high resolution aneurysm mesh (28,368 triangles). (b): the one-sided Hausdorff distance error map on a mesh of 772 shells generated with our method.

Coupling between tetrahedra and shells for advanced modelling. Structures in human body can be either solid (brain, liver, prostate etc.) or hollow (colon, blood vessels, stomach etc.). However knowing how to model the two kind of structures is not sufficient to reach a high degree of accuracy, real life situations are more complex. As an example, the external surface of the liver is covered by a layer of cells called Glisson's capsule. Its interaction with the liver plays an important role into the overall structure's mechanical behaviour. Therefore considering the interaction between solid and hollow objects is as crucial as modelling the two structures separately.

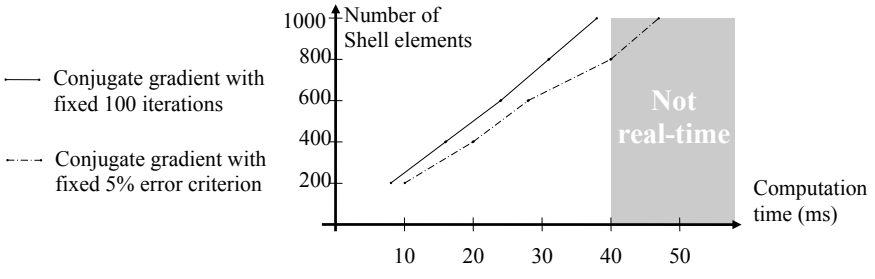


Fig. 6. Computation time on meshes of 200, 400, 600, 800 and 1000 elements

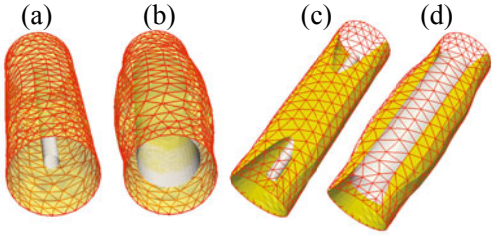


Fig. 7. Simulation of an angioplasty procedure. (a, c): A collapsed stent is inserted into the blood vessel. (b, d): The stent is crushing the fatty deposits which creates a pressure onto the interior wall and widens the blood vessel.

An example of medical procedure to illustrate this point even further is angioplasty. Angioplasty is the technique of mechanically widening a narrowed or obstructed blood vessel, typically as a result of atherosclerosis. An empty and collapsed balloon on a guide wire is passed into the narrowed locations and then inflated to a fixed size. The balloon crushes the fatty deposits, so opening up the blood vessel to improved flow. As a proof of concept we tried to simulate an angioplasty (Fig. 7). The blood vessel is modelled using the shell FEM formulation described in this paper and the fatty deposits are simulated with a tetrahedral FEM method and are fixed to the interior wall of the blood vessel.

When the balloon inflates it crushes the deposits and they then apply a pressure onto the curved surfaces of shells modelling the interior wall. The forces are then distributed onto the mechanical nodes of the blood vessel mesh as detailed in section 2, which widens the blood vessel as expected.

5 Conclusion

We propose a framework for real-time modelling of thin anatomical structures. The novelty of our method relies on the combination of a shell finite element formulation and a geometric surface reconstruction both based on the same polynomial interpolation function used to describe the surface of shells. We also show how contacts and interactions with the curved surfaces of shells can be handled using the same function. The efficiency of the method is illustrated through shell-based reconstruction and real-time simulation of the deformations of various anatomical structures. We also present preliminary results on the simulation of an angioplasty procedure.

References

1. 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(2), 121–134 (2007)
2. Delingette, H.: Biquadratic and Quadratic Springs for Modeling St Venant Kirchhoff Materials. In: Bello, F., Edwards, E. (eds.) ISBMS 2008. LNCS, vol. 5104, pp. 40–48. Springer, Heidelberg (2008)
3. de Visser, H., Passenger, J., Conlan, D., Russ, C., Hellier, D., Cheng, M., Acosta, O., Ourselin, S., Salvado, O.: Developing a next generation colonoscopy simulator. Int. J. Image Graphics 10(2), 203–217 (2010)
4. Liu, G., Quek, S.: Finite Element Method: A Practical Course. Butterworth (2003)
5. Nedel, L.P., Thalmann, D.: Real time muscle deformations using mass-spring systems. In: Proceedings of Computer Graphics International, p. 156 (1998)
6. Hammer, P.E., Perrinb, D.P., del Nidob, P.J., Howe, R.D.: Image-based mass-spring model of mitral valve closure for surgical planning. In: Proceedings of SPIE Medical Imaging, vol. 6918 (2008)
7. Choi, M.G., Woo, S.Y., Ko, H.S.: Real-time simulation of thin shells. In: ACM SIGGRAPH/Eurographics symposium on Computer animation, pp. 349–354 (2007)
8. Comas, O., Cotin, S., Duriez, C.: A shell model for real-time simulation of intraocular implant deployment. In: Bello, F., Cotin, S. (eds.) Biomedical Simulation. LNCS, vol. 5958, pp. 160–170. Springer, Heidelberg (2010)
9. Kolb, A., Pottmann, H., Peter Seidel, H.: Fair surface reconstruction using quadratic functionals. Computer Graphics Forum 14(3), 469–479 (1995)
10. Tang, C.K., Medioni, G.: Robust estimation of curvature information from noisy 3d data for shape description. In: Proceedings of IEEE ICCV (1999)
11. Allard, J., Cotin, S., Faure, F., Bensoussan, P.-J., Poyer, F., Duriez, C., Delingette, H., Grisoni, L.: SOFA - an Open Source Framework for Medical Simulation. In: Medicine Meets Virtual Reality, pp. 13–18 (2007)

12. Przemieniecki, J.: Theory of matrix structural analysis. McGraw-Hill, New York (1985)
13. Klein, R., Liebich, G., Straßer, W.: Mesh reduction with error control. In: Visualization 96, pp. 311–318. ACM, New York (1996)
14. CNR, V.C.L.I.: Meshlab, http://meshlab.sourceforge.net/
15. Saupin, G., Duriez, C., Grisoni, L.: Embedded multigrid approach for real-time volumetric deformation. 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. 149–159. Springer, Heidelberg (2007)

Personalization of Cubic Hermite Meshes for Efficient Biomechanical Simulations

Pablo Lamata[1], Steven Niederer[1], David Barber[2], David Norsletten[1], Jack Lee[1], Rod Hose[2], and Nic Smith[1]

[1] Computing Laboratory, University of Oxford, UK
{pablo.lamata,steven.niederer,david.norsletten,
jack.lee,nic.smith}@comlab.ox.ac.uk
[2] Department of Cardiovascular Science, University of Sheffield, UK
{d.barber,d.r.hose}@sheffield.ac.uk

Abstract. Cubic Hermite meshes provide an efficient representation of anatomy, and are useful for simulating soft tissue mechanics. However, their personalization can be a complex, time consuming and labour-intensive process. This paper presents a method based on image registration and using an existing template for deriving a patient-specific cubic Hermite mesh. Its key contribution is a solution to customise a Hermite continuous description of a shape with the use of a discrete warping field. Fitting accuracy is first tested and quantified against an analytical ground truth solution. To then demonstrate its clinical utility, a generic cubic Hermite heart ventricular model is personalized to the anatomy of a patient, and its mechanical stability is successfully tested. The method achieves an easy, fast and accurate personalization of cubic Hermite meshes, constituting a crucial step for the clinical adoption of physiological simulations.

1 Introduction

Computational physiology provides tools to quantitatively describe physiological behaviour across a range of time scales and anatomical levels using mathematical and computational models [1,2]. The heart is arguably the most advanced current exemplar of this approach [3,4], and ongoing developments now have the potential to provide a significant impact in the management of cardiovascular diseases. One of the key challenges in fulfilling this potential is the personalization of models to represent the clinical status of a patient. This work focuses on the efficient and automated development of patient-specific geometrical description of organs for biomechanical simulations.

A computational model requires the geometrical description of the solution domain where material constitutive equations are solved. The most popular choice is linearly interpolated tetrahedral meshes, mainly due to its conceptual simplicity and availability of tools for an automatic mesh generation [5]. Nevertheless, they introduce significant numerical error in the solution of the displacements in the soft tissue deformation problem [6]. Alternatively, cubic Hermite meshes are

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 380–387, 2010.

an efficient representation of the geometrical state of an organ, and a more suitable choice for biomechanical simulations compared to tetrahedral meshes [6]. Another important requirement is mesh robustness, related to the convergence of simulation results in changing physiological conditions. For these reasons, cubic Hermite meshes are a popular choice for the simulation of heart mechanical deformations [7,8,9]. Nevertheless, construction of these meshes can be a complex, time consuming and labour-intense process. There is thus a need for a fast, accurate, robust and easy to use cubic Hermite personalization method.

There are two broad approaches for the personalization of geometrical meshes: construction from segmented images [5,10] or customization from an existing mesh model [11,10]. Whereas the literature for linear meshes is extensive [5], its translation to Hermite meshes is not straightforward. The change of interpolation scheme, from linear to Hermite functions, requires a completely different meshing strategy, like the adaptation of the Iterative Closest Point proposed in [10]. In a Hermite mesh shape is interpolated, not only from the 3D Cartesian coordinates of nodes, but also from the derivatives of shape versus local finite element (FE) coordinates. This enables a compact representation, but results in a complex mesh construction and customization. This article presents an image registration based solution for cubic Hermite mesh personalization.

2 Material and Methods

The proposed personalization method combines a fast binary image registration with a cubic Hermite warping technique. A schematic illustration of the complete process is provided in Fig.1.

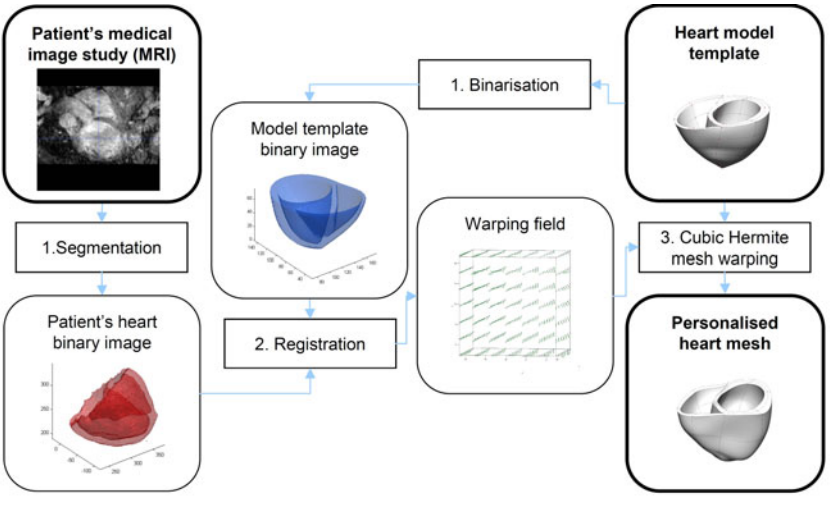


Fig. 1. Dataflow designed to generate patient specific cubic Hermite meshes

2.1 Image Registration

A binary image registration technique using a fast optical flow algorithm proposed in [11] is chosen for its robustness, accuracy and computational efficiency. It requires a preliminary segmentation of the patient's anatomy, which in the case of heart ventricles from static MRI or CT is a quite mature field of research. Discrete values of the warping field between the two binary images are defined at nodes of a regular hexahedral grid superimposed on the binary image of the template shape. The two main parameters of the registration technique are the spacing between nodes D and the smoothing weighting factor λ (based on a Tikhonov regularisation of the linear least squares problem). Registration is initialised by aligning the principal axis of the shapes after a Principal Component Analysis of the 3D coordinates of each shape.

2.2 Cubic Hermite Mesh Warping

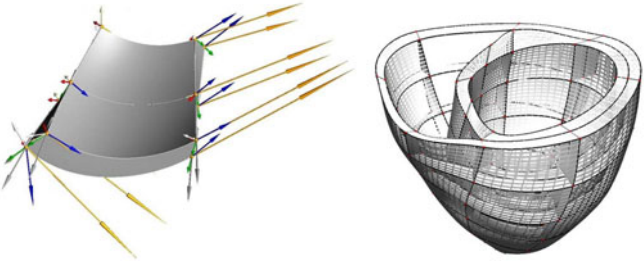
A cubic Hermite mesh is a set of 3D FE that uses Hermite interpolation functions. Mesh nodes have both coordinate values and derivatives (single, double and triple cross derivatives) in order to encode a C^1 shape. Let $u(\xi_1, \xi_2, \xi_3)$ be a shape defined in the Cartesian space as a function of the material coordinates ξ. The four 1D Hermite interpolation basis functions are described in (1), and interpolation in a line element $u(\xi)$ is given by a linear combination of these four basis functions (2). This scheme can be extended to 3D as illustrated in Fig.2, $u(\xi_1, \xi_2, \xi_3)$ is then interpolated from a total of 192 variables in each element.

$$\begin{aligned} \psi_1^0(\xi) = 1 - 3\xi^2 + 2\xi^3;\ \ \psi_2^0(\xi) = \xi^2(3 - 2\xi) \\ \psi_1^1(\xi) = \xi(\xi - 1)^2;\ \ \psi_2^1(\xi) = \xi^2(\xi - 1) \end{aligned} \tag{1}$$

$$u(\xi) = \psi_1^0(\xi)u_1 + \psi_2^0(\xi)u_2 + \psi_1^1(\xi)du/d\xi|_1 + \psi_2^1(\xi)du/d\xi|_2 \tag{2}$$

Warping a cubic Hermite mesh therefore implies calculation of a deformed state of its three coordinate fields (x, y, z). The total number of degrees of freedom is therefore $N_{dof} = 3 \times 8 \times n = 24n$, being n the number of nodes of the mesh. Note that this number will be slightly smaller if the mesh has collapsed elements, and slightly bigger if there are discontinuities modelled with different versions of node values. For further details about these meshes see [10].

The solution for the warping of a cubic Hermite mesh is the core contribution of this work. This third step in Fig.1 calculates the optimal description of a warped shape with Hermite interpolation functions. This is built on three basic concepts. First, it is important to realise that the warping of a FE requires knowing the warping field throughout the complete domain, and not only in the local neighbourhood of mesh nodes. Second, the solution uses a FEM technique for finding an optimal representation of a field in a domain, i.e. it uses a variational formulation of the problem based on a dot product of functions. And third, a numerical technique is required to handle continuous domains with computer discrete representations, i.e. a numerical integration method is used to calculate the dot product of functions.



(a) (b)

Fig. 2. Cubic Hermite meshes. (a) A two element mesh, showing the vectors corresponding to the 7 derivatives at each node; (b) template of truncated left and right ventricles of the heart. This complex shape is represented with only 112 cubic elements (cubes with black lines) and 183 nodes.

The central idea is that warping of a cubic Hermite mesh is an addition of a FE description of a warping field. It requires finding the adequate FE description, with a set of N_{dof} variables as described before, of the warping field. And this description is then added to the corresponding N_{dof} nodal variables describing the 3 coordinate fields, i.e. the shape. Let us define a variational problem for finding the C^1 continuous function g that approximates w, one of the three components of the warping field W in the domain defined by the cubic Hermite template Θ_T. Let g be formulated as a linear combination of the set of $N_{dof}/3$ basis functions ϕ_T of the mesh. Note that each ϕ_T is a combination of 3D Hermite basis from adjacent elements in order to enforce the continuity of the function and its derivatives. Let us introduce the definition of a dot product of functions:

$$(g, \phi_j) = \int_{\Theta} g \cdot \phi_j = \int_{\Theta} \sum_i c_i \phi_i \cdot \phi_j = \sum_i c_i \int_{\Theta} \phi_i \cdot \phi_j \tag{3}$$

Finding g becomes the problem of finding the c_i coefficients (or nodal values) which satisfy a set of $N_{dof}/3$ equations, one for each basis ϕ_T:

$$(g, \phi_j) = (w, \phi_j) \forall \phi_j \tag{4}$$

$$\sum_i c_i \int_{\Theta_T} \phi_i \cdot \phi_j = \int_{\Theta_T} w \cdot \phi_j \tag{5}$$

$$Mx = b \tag{6}$$

where vector x is the set of $N_{dof}/3$ coefficients c_i, and integrals required to calculate matrix M and vector b are computed using Gaussian Quadrature. Calculi of vector b uses order 4, and cubic interpolation of the warping field is used to calculate the data (deformation field) at Gauss Points. The linear system of the fitting process requires the solution of a sparse matrix system. This matrix is symmetric and positive definite, allowing the use of fast, low-memory solvers such as Conjugate Gradients.

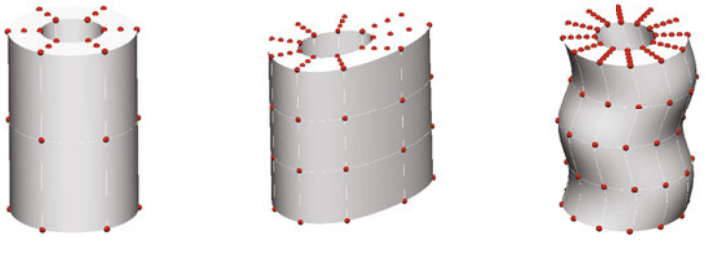
3 Results

3.1 Analytical Workbench for Accuracy and Robustness Analysis

Accuracy and sensitivity to parameters are analysed using a virtual workbench with known analytical solutions. Three experiments study the accuracy of (1) proposed warping scheme, step 3 in Fig.1, (2) the binary image registration, step 2 in Fig.1, and (3) the concatenation of the two processes. Template cylinders are built with dimensions 30, 10 and 5mm in length, outer and internal radius, and with 24, 81 and 192 elements. Warped versions of these cylinders are generated under two known warping fields W_1 and W_2, see Fig.3. Shape error is calculated in each element as the integral of the RMS error between warped coordinates field and their ground truth. An order 5 Gaussian Quadrature volume-weighted integration of error is used, which is independent of the mesh discretisation resolution and topology.

The warping step is analysed by using a perfect solution of the registration step, obtained by sampling the ground truth warping fields. This experiment is repeated for W_1 and W_2, for the three cylinder mesh resolutions and for a range of node spacings in the discretisation of the warping field (S_W, with 9 values from 1 to 5 mm in steps of 0.5mm). Average shape error is $5.5e^{-3}$ and $5.5e^{-2}mm^2$ for W_1 and W_2 respectively, and the dependence with the two factors (mesh resolution and S_W) is illustrated in Fig.4.

The registration step is analysed by comparison of obtained warping fields with their analytical expression. Binary images are generated from the Ground Truth shapes. Registration is performed for each W_1 and W_2, for a total of 49 binary image resolutions (from 0.2mm to 5mm in steps of 0.1mm), and for 9 values of node spacing (D from 2 to 10 voxels). The smoothing coefficient λ is set automatically by an empirical theorem proposed in [11]. Average registration volume-weighted error is 1.86 and $0.69mm$ RMS for W_1 and W_2 respectively, results are shown in Fig.5.



(a) *Template* (b) $W_1 = (x^3, x^3, 1)$ (c) $W_2 = (sin(z), 0, 0)$

Fig. 3. GroundTruth. All three shapes (template, customised by W_1 and customised by W_2) are created with three resolutions (24, 81 and 192 elements).

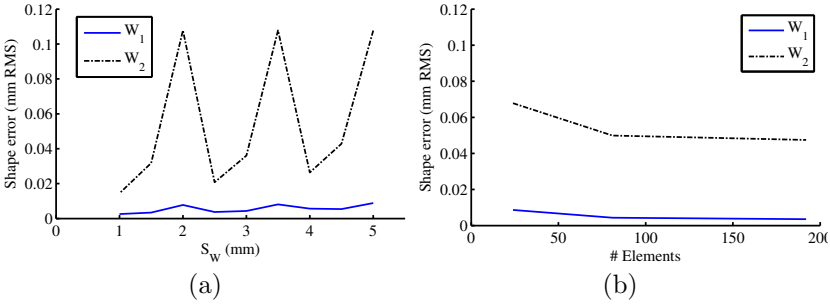


Fig. 4. Error introduced in the warping of cubic Hermite meshes using a discrete version of the ideal warping fields W_1 and W_2. (a) Dependence on sampling resolution, (S_W), averaging results with three mesh resolutions. (b) Dependence on mesh resolution, averaging results with nine sampling resolutions.

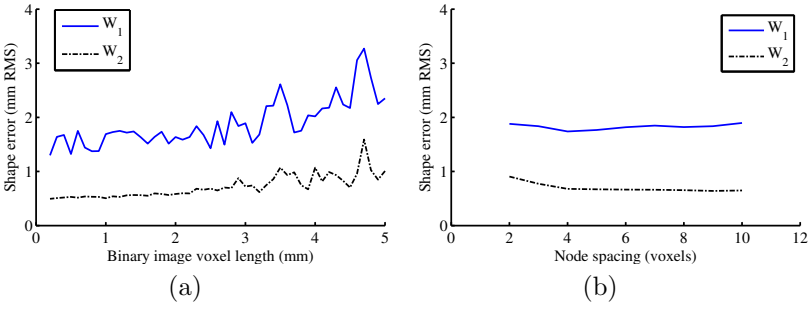


Fig. 5. Registration error (volume-weighted integral) for warping fields W_1 and W_2. (a) Dependence on image resolution, averaging results with 9 node spacings. (b) Dependence on node spacing, averaging results with 49 image resolutions.

The previous two steps are concatenated, and template cylinder meshes are warped with the result of binary image registrations. The mesh resolution of 81 elements is chosen for this experiment, since higher resolutions did not introduce any significant improvement in accuracy, see Fig.4b. The analysis is repeated for W_1 and W_2, for all 49 binary image resolutions (from 0.2mm to 5mm in steps of 0.1mm), and for 9 node spacing (from 2 to 10 voxels). The average shape error for W_1 and W_2 is 1.90 and 0.702mm RMS respectively, an increment of roughly a 1.5% with respect to the registration error. The dependence on binary resolution and node spacing is the same as reported for the registration step, see Fig.5.

3.2 Clinical Case

A patient specific cubic Hermite mesh of heart ventricles is constructed following the process of Fig.1. The template mesh chosen is the result of fitting a mesh to the anatomy of a first patient following the methodology described in [10] and

illustrated in Fig.2. A T1 MRI heart study of a second patient (0.88x0.88x0.75 mm voxel resolution) is manually segmented and truncated just underneath of the opening of valve planes after vertical alignment. The agreement between segmentation and resulting cubic Hermite mesh is shown in Fig.6, and the average distance between the two surfaces is measured as 1.32mm. The process from image segmentation to mesh fitting for this biventricular dataset of 90^3 voxels requires about one minute. Mechanical stability of personalised geometry is successfully tested by simulating a passive inflation and an isochronous active contraction of the heart following the methods described in [7].

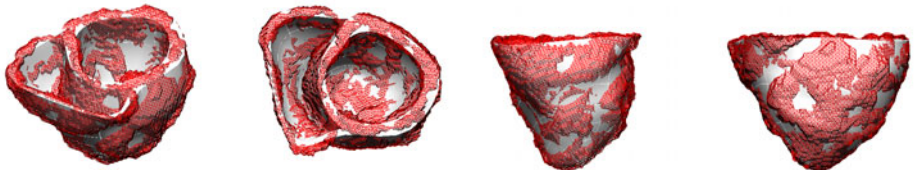


Fig. 6. Shape personalization result. Comparison of the isosurface of the binary manual segmentation (red wireframe) to the Cubic Hermite mesh (white solid).

4 Discussion

Cubic Hermite mesh warping requires the addition of an adequate representation of the warping field in nodal values and derivatives. The proposed solution finds an optimal description of this field, leading to reasonable accurate results.

Results show the importance of the registration step, which is limited by image resolution. Interpolation errors also become significant when the topology of the mesh is not able to represent the warped shape. These cases require an interpolation order higher than cubic or higher element refinement, such as case W_2. An inherent limitation of proposed approach is that binary registration only aligns the surfaces of models, interpolating the warping inside. This is a valid approach for vessels and computational fluid dynamics [11], and preliminary results suggest that this is also adequate for the flat walls of the heart.

The proposed method is its simpler and more robust compared to the state of the art alternatives, the "host mesh" technique [10] or the mesh generation by fitting Hermite surfaces from a linear scaffold [10]. Proposed method is fast and requires minimal user interaction. In comparison, the two alternative methods greatly depend on users expertise, and it can take hours of manual interaction and fitting to get satisfactory results. A second advantage is that, because it uses a voxelized description of the shape and not a selection of pairs of control points and degrees of freedom as required in a "host mesh" technique, it is immune to subjective and sometimes arbitrary selection of parameters. Finally, warping an existing high quality template mesh under a smoothness constraint is a reasonable warranty of simulation stability in the resulting mesh. This does not occur using a mesh generation from an arbitrary initial linear mesh [10]. Further experiments with higher number of cases are nevertheless required to

generalise and accept these desirable properties. Future works will also address the characterization and development of a metric of mechanical stability of cubic Hermite meshes, a metric to be optimised during the personalization process.

5 Conclusion

Proposed method achieves an easy, fast and accurate personalization of cubic Hermite meshes, constituting a crucial step for the clinical adoption of biomechanical physiological simulations.

References

1. Hunter, P.J., Crampin, E.J., Nielsen, P.M.F.: Bioinformatics, multiscale modeling and the IUPS Physiome Project. Brief Bioinform 9(4), 333–343 (2008)
2. Lee, J., Niederer, S., Nordsletten, D., Le Grice, I., Smail, B., Kay, D., Smith, N.: Coupling contraction, excitation, ventricular and coronary blood flow across scale and physics in the heart. Phil. Trans. R Soc. A 367(1896), 2311–2331 (2009)
3. Smith, N.P., Nickerson, D.P., Crampin, E.J., Hunter, P.J.: Multiscale computational modelling of the heart. Acta Numerica 13(1), 371–431 (2004)
4. Bassingthwaighte, J., Hunter, P., Noble, D.: The Cardiac Physiome: perspectives for the future. Experimental Physiology 94(5), 597–605 (2009)
5. Löhner, R.: Automatic unstructured grid generators. Finite Elem. Anal. Des. 25(1-2), 111–134 (1997)
6. Pathmanathan, P., Whiteley, J.P., Gavaghan, D.J.: A comparison of numerical methods used for finite element modelling of soft tissue deformation. The Journal of Strain Analysis for Engineering Design 44(5), 391–406 (2009)
7. Niederer, S.A., Smith, N.P.: The role of the Frank Starling law in the transduction of cellular work to whole organ pump function: A computational modeling analysis. PLoS Comput. Biol. 5(4), e1000371 (2009)
8. Kerckhoffs, R.C., McCulloch, A.D., Omens, J.H., Mulligan, L.J.: Effects of biventricular pacing and scar size in a computational model of the failing heart with left bundle branch block. Medical Image Analysis 13(2), 362–369 (2009)
9. Wang, V.Y., Lam, H., Ennis, D.B., Cowan, B.R., Young, A.A., Nash, M.P.: Modelling passive diastolic mechanics with quantitative mri of cardiac structure and function. Medical Image Analysis 13(5), 773–784 (2009)
10. Fernandez, J., Mithraratne, P., Thrupp, S., Tawhai, M., Hunter, P.: Anatomically based geometric modelling of the musculo-skeletal system and other organs. Biomechanics and Modeling in Mechanobiology 2, 139–155 (2004)
11. Barber, D., Oubel, E., Frangi, A., Hose, D.: Efficient computational fluid dynamics mesh generation by image registration. MedIA 11(6), 648–662 (2007)

Real-Time Surgical Simulation Using Reduced Order Finite Element Analysis

Zeike A. Taylor[1,2], Stuart Crozier[1], and Sébastien Ourselin[2]

[1] MedTeQ Centre, School of Information Technology & Electrical Engineering, The University of Queensland, Brisbane, QLD, 4072, Australia
[2] Centre for Medical Image Computing, University College London, Gower St, London, WC1E 6BT, UK
ztaylor@itee.uq.edu.au

Abstract. Reduced order modelling, in which a full system response is projected onto a subspace of lower dimensionality, has been used previously to accelerate finite element solution schemes by reducing the size of the involved linear systems. In the present work we take advantage of a secondary effect of such reduction for explicit analyses, namely that the stable integration time step is increased far beyond that of the full system. This phenomenon alleviates one of the principal drawbacks of explicit methods, compared with implicit schemes. We present an explicit finite element scheme in which time integration is performed in a reduced basis. The computational benefits of the procedure within a GPU-based execution framework are examined, and an assessment of the errors introduced is given. Speedups approaching an order of magnitude are feasible, without introduction of prohibitive errors, and without hardware modifications. The procedure may have applications in medical image-guidance problems in which both speed and accuracy are vital.

1 Introduction

In the last decade significant effort has been devoted to use of computational biomechanics for enhancing the utility of medical images in the operating theatre. A typical scenario is the use of information-rich pre-operative images to guide interventions. Such images may, for example, delineate target pathologies or vasculature and other vital structures. However, if the relevant biological structures change shape between imaging and intervention, guidance may be confounded unless the deformation can be reliably compensated for. Since soft tissues are, of course, physical structures, their deformation must conform to physical laws, and this has motivated use of biomechanics to this end [1].

A common issue is the stringent time constraint on simulations performed intra-operatively, which competes with requirements for accuracy and reliability in the solution. Continuum mechanics formalism, combined with finite element solution, is the most common framework. However, despite the maturity of nonlinear formulations in the biomechanics community, these have seldom been used, largely due to the mentioned time constraints.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 388–395, 2010.

Recently, explicit dynamic algorithms have been shown to be advantageous in this area [12,13], but a well-known drawback of such methods is the small time steps required for numerical stability. By virtue of its dependence on material stiffness, the critical step size Δt_{cr} is much larger for models of very soft tissues (such as organs) than of common engineering materials, but is still many times smaller than that required in implicit analyses. In the present work, we examine the use of reduced order models (ROMs) to alleviate some of the computational load in explicit analyses, and thereby improve their utility.

The main idea behind ROMs is projection of the full model response onto a lower dimensional *reduced basis*. In a displacement-based finite element context, the complete displacement field is approximated by a set of generalised variables of much lower dimension. In most developments in this area [6,8], the key benefit is the reduction in the size of the linear system to be solved at each step, since mass and stiffness matrices assume the dimensionality of the reduced system. However, in explicit analyses, these large matrices are removed, and the utility of the ROMs is less obvious. Our work makes use of a secondary effect for explicit methods: by performing the time integration in the reduced basis, a much larger solution time step may be used [6]. This phenomenon does not appear to have been exploited since its discovery by Krysl and co-workers, despite its potential benefits.

In the following, we describe an explicit solution method which incorporates integration in the reduced basis, as described. The computational benefits of the procedure within a GPU-based execution framework are examined, subsequently, along with the errors incurred.

2 A Reduced Order Explicit Dynamic FE Algorithm

In the following we draw on the work of Taylor et al. [12,13], in which an explicit dynamic FE algorithm was employed. We refer the reader to these publications for further algorithm details.

2.1 Equilibrium Equations

We are concerned with solution of the standard equilibrium equations for a nonlinear, dynamic, damped finite element model:

$$\mathbf{M}\ddot{\mathbf{U}} + \alpha\mathbf{M}\dot{\mathbf{U}} + \mathbf{K}(\mathbf{U})\mathbf{U} = \mathbf{F}^{\mathrm{ext}}, \tag{1}$$

where $\mathbf{M}$ is the (diagonalised) mass matrix, α is a damping coefficient, $\mathbf{K}(\mathbf{U})$ is the stiffness matrix, $\mathbf{F}^{\mathrm{ext}}$ is a vector of external loads, and $\mathbf{U}$ is a vector of nodal displacements. When explicit time integration is used, the stiffness term $\mathbf{K}(\mathbf{U})\mathbf{U}$ may be evaluated element-wise and converted into an equivalent vector of internal forces $\mathbf{F}^{\mathrm{int}}$ (see [12] for details). Thus, at time increment n, the equilibrium equations read

$$\mathbf{M}\ddot{\mathbf{U}}_n + \alpha\mathbf{M}\dot{\mathbf{U}}_n = \mathbf{F}_n^{\mathrm{eff}}, \tag{2}$$

where $\mathbf{F}_n^{\mathrm{eff}} = \mathbf{F}_n^{\mathrm{ext}} - \mathbf{F}_n^{\mathrm{int}}$ is called the effective load vector.

2.2 Generalised Displacements and the Reduced Basis

The system defined by Eqn. (2), referred to as the *full* system, has dimension N, and its response is encapsulated by the nodal displacements $\mathbf{U}$. For the present case of a displacement-based finite element method in 3D, we have $N = 3N_{\text{nodes}}$, where N_{nodes} is the number of nodes. The key idea of reduced order modelling in a finite element context is to approximate the full response with a set of *generalised displacements* $\mathbf{P}$ of much lower dimension M:

$$\mathbf{U} = \mathbf{\Phi}\mathbf{P}, \tag{3}$$

where $\mathbf{\Phi}$ is referred to as the *reduced basis*. $\mathbf{\Phi}$ is time-independent, thus: $\dot{\mathbf{U}} = \mathbf{\Phi}\dot{\mathbf{P}}$, $\ddot{\mathbf{U}} = \mathbf{\Phi}\ddot{\mathbf{P}}$. Substituting these into (2) and projecting on $\mathbf{\Phi}$ yields

$$\hat{\mathbf{M}}\ddot{\mathbf{P}}_n + \alpha\hat{\mathbf{M}}\dot{\mathbf{P}}_n = \hat{\mathbf{F}}^{\text{eff}}_n, \tag{4}$$

in which we have introduced a reduced mass matrix $\hat{\mathbf{M}} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{M}\mathbf{\Phi}$ and reduced effective load $\hat{\mathbf{F}}^{\text{eff}} = \mathbf{\Phi}^{\mathrm{T}}\mathbf{F}^{\text{eff}}$. Eqn. (4) constitutes an equivalent equilibrium equation for the reduced system.

2.3 Computing the Reduced Basis

For a full-order dynamic FE model subjected to some loading, we define a matrix $\bar{\mathbf{U}}$ of snapshots:

$$\bar{\mathbf{U}} = \left[\, \mathbf{u}_1 - \bar{\mathbf{u}}\ \mathbf{u}_2 - \bar{\mathbf{u}} \ldots \mathbf{u}_S - \bar{\mathbf{u}} \,\right], \tag{5}$$

where $\mathbf{u}_i$ $(i = 1, \ldots, S)$ are N-vectors of nodal displacements captured at S different time points, and $\bar{\mathbf{u}} = (\sum_{i=1}^{S} \mathbf{u}_i)/S$ is their mean. $\mathbf{u}_i$ are referred to as snapshots, and we call the collection of them the training set for the reduced model. We seek an optimal M-dimensional approximation for the response history $\bar{\mathbf{U}}$, with $M \ll N$. It may be shown that this is given by the eigenvectors $\boldsymbol{\phi}_i$ $(i = 1, \ldots, M)$[1] corresponding to the M largest eigenvalues of the covariance matrix $\mathbf{C}_d = \bar{\mathbf{U}}\bar{\mathbf{U}}^{\mathrm{T}}/M$ [6,8]. The desired reduced basis $\mathbf{\Phi}$ is then given by

$$\mathbf{\Phi} = \left[\, \boldsymbol{\phi}_1\ \boldsymbol{\phi}_2 \ldots \boldsymbol{\phi}_M \,\right]. \tag{6}$$

The covariance matrix $\mathbf{C}_d$ is $N \times N$. Hence, for practical problems, obtaining the required eigenvectors is computationally intensive. For $S < N$ (which is usual) a more efficient approach exists. Begin by computing the first M eigenvectors $\boldsymbol{\psi}_i$ of the smaller matrix $\mathbf{C}_s = \bar{\mathbf{U}}^{\mathrm{T}}\bar{\mathbf{U}}/M$. The corresponding eigenvalues are denoted λ_i. Scale each eigenvector according to $\bar{\boldsymbol{\psi}}_i = \boldsymbol{\psi}_i/\sqrt{\lambda_i}$. The desired basis is then recovered by projecting back on the matrix of snapshots: $\mathbf{\Phi} = \bar{\mathbf{U}}\bar{\mathbf{\Psi}}$, where $\bar{\mathbf{\Psi}}$ is a matrix whose columns are the scaled vectors $\bar{\boldsymbol{\psi}}_i$.

In the study in [6], the reduced models were trained using snapshots from a single load case. Models so constructed are able to reproduce very well the full response to this load, but are less accurate when subjected to different loads. We have found that the robustness of the reduced models may be greatly improved by including snapshots from several different load cases in the training set.

[1] We refer to these as the *first* M eigenvectors.

2.4 Explicit Integration in the Reduced Basis

In an incremental nonlinear analysis, we require a procedure for computing the new model configuration at each step. Assume, temporarily, that the generalised displacements $\mathbf{P}_{n-1}$ and $\mathbf{P}_n$ for the previous and current steps, respectively, are known, and the reduced effective load $\hat{\mathbf{F}}_n^{\text{eff}}$ has been computed. We define discrete approximations for current generalised accelerations and velocities, $\ddot{\mathbf{P}}_n = (\mathbf{P}_{n-1} - 2\mathbf{P}_n + \mathbf{P}_{n+1})/\Delta t^2$ and $\dot{\mathbf{P}}_n = (\mathbf{P}_{n+1} - \mathbf{P}_{n-1})/2\Delta t$, respectively, in which Δt is the solution time step. Substituting these in the equilibrium equation (4) and rearranging yields an expression for the next set of generalised displacements:

$$\mathbf{P}_{n+1} = \gamma_1 \hat{\mathbf{M}}^{-1} \hat{\mathbf{F}}_n^{\text{eff}} + \gamma_2 \mathbf{P}_n + \gamma_3 \mathbf{P}_{n-1}, \tag{7}$$

where $\gamma_1 = 2\Delta t^2/(\alpha \Delta t + 2)$, $\gamma_2 = 2\gamma_1/\Delta t^2$, and $\gamma_3 = 1 - \gamma_2$ are constants. Then, the full displacements may be recovered by multiplying through by $\mathbf{\Phi}$:

$$\begin{aligned} \mathbf{U}_{n+1} &= \mathbf{\Phi} \mathbf{P}_{n+1} \\ &= \gamma_1 \mathbf{\Phi} \hat{\mathbf{M}}^{-1} \hat{\mathbf{F}}_n^{\text{eff}} + \gamma_2 \mathbf{\Phi} \mathbf{P}_n + \gamma_3 \mathbf{\Phi} \mathbf{P}_{n-1} \\ &= \gamma_1 \mathbf{\Phi} \hat{\mathbf{M}}^{-1} \mathbf{\Phi}^{\mathrm{T}} \mathbf{F}_n^{\text{eff}} + \gamma_2 \mathbf{U}_n + \gamma_3 \mathbf{U}_{n-1}. \end{aligned} \tag{8}$$

$\hat{\mathbf{M}}$ is a dense matrix, but of much smaller size ($M \times M$ where $M \ll N$) than the full mass matrix $\mathbf{M}$. In practice its inverse is easily precomputed and stored, and occupies minimal storage. A more pertinent issue from the point of view of parallelisation is that the displacement update is no longer a simple vector equation. This is addressed in Sect. 3.

The new displacement update procedure (8) consists of projecting the effective load on the reduced basis, evaluating a generalised acceleration term, and projecting back on the full system. As mentioned, the principal benefit of doing so is that it *increases the stable time step far beyond that of the full system.* While the precise mechanism for this has not been established, it is likely that by evaluating the accelerations in the reduced basis we remove many of the higher frequency modes of the full system, thus increasing the effective minimum free-vibrational period on which the critical limit depends. Many of these frequencies result from discretisation artifacts and do not contribute meaningful information to the solution. Many others, though physical, may be neglected also without significantly altering the solution, as will be seen.

2.5 Consistency of the Nonlinear Formulation

It is important to note that the only modification of the algorithm in [12,13] is in the time integration procedure; the evaluation of stresses at integration points and their integration to obtain internal forces $\mathbf{F}^{\text{int}}$ are unchanged. Thus, assuming Eqn. (8) provides a sufficiently accurate approximation for the full displacements, the large deformation consistency of the nonlinear formulation is intact.

3 GPU Implementation Using CUDA

Using the CUDA API [9], the proposed algorithm was implemented for GPU execution. As mentioned, the introduction of the matrix multiplications in the displacement update (8) complicates parallel execution. Instead of the single computation kernel for this process used in [12,13], we require several. Execution is divided into precomputation and time-loop phases. The precomputation phase is as described in [12,13] except for the computation of the inverse reduced mass matrix $\hat{\mathbf{M}}^{-1}$. At each step in the time-loop the execution procedure is now:

1. Compute internal forces $\mathbf{F}_n^{\text{int}}$ **[GPU]**
2. Compute effective loads $\mathbf{F}_n^{\text{eff}} = \mathbf{F}_n^{\text{ext}} - \mathbf{F}_n^{\text{int}}$ **[GPU]**
3. Transfer $\mathbf{F}_n^{\text{eff}}$ to host
4. Compute reduced effective loads $\hat{\mathbf{F}}_n^{\text{eff}} = \mathbf{\Phi}^{\text{T}}\mathbf{F}_n^{\text{eff}}$ **[CPU]**
5. Compute $\hat{\mathbf{F}}_n^{\text{eff}} = \hat{\mathbf{M}}^{-1}\hat{\mathbf{F}}_n^{\text{eff}}$ **[CPU]**
6. Compute $\mathbf{U}_{n+1} = \mathbf{\Phi}\hat{\mathbf{F}}_n^{\text{eff}}$ **[CPU]**
7. Transfer $\mathbf{U}_{n+1}$ to device
8. Compute $\mathbf{U}_{n+1} = \gamma_1\mathbf{U}_{n+1} + \gamma_2\mathbf{U}_n + \gamma_3\mathbf{U}_{n-1}$ **[GPU]**

Bold, bracketed comments indicate use of CPU or GPU execution. Notably, steps 4-6 are executed on the CPU, which, owing to the small sizes of the involved quantities, we found to be faster than invoking additional CUDA kernel launches. Steps 1, 2, and 8 are implemented as CUDA kernel functions. Thus, the new algorithm requires three kernel launches (plus two data transfers and some additional CPU computations), compared with two for the full model implementation [12,13].

4 Assessment Using a Brain Shift Model

We assessed the robustness and computational performance of the presented algorithm using a model of the well-known brain shift problem [4]. Many methods for compensating for the brain motion using finite element analysis have been proposed [3,7,10,14]. We studied a synthetic scenario in which the skull is imagined to be opened adjacent to the right frontal lobe, and brain motion is induced in this region.

The brain tissue was modelled as a neo-Hookean hyperelastic material with shear modulus $\mu = 1$kPa and bulk modulus $\kappa = 50$kPa [14]. A model comprising 7480 nodes and 39 323 tetrahedral elements (non-locking Averaged Nodal Pressure formulation was used [5]) was employed. Loads were applied over 1sec of simulated time (sufficiently slow to obtain an approximately stable static solution) to a group of nodes on the right frontal lobe, and surface nodes on the opposite side of the model (left, posterior) were assumed to be in contact with the skull and were fixed. In the present context, only the final deformed configuration of the brain is of interest.

We emphasise that this study is not based on a particular clinical case, and is intended solely to enable comparison of the reduced order modelling approach with the well-established regular approach for a clinically relevant scenario. That is, we seek to establish whether the same results are obtained with each method.

4.1 A Priori Construction of the Reduced Basis

Reduced bases $\mathbf{\Phi}$ were constructed by compiling full model response snapshots from five independent load cases. The applied loads in each case were of similar magnitude, but varying direction, with the directions covering an angular range of 40°. 100 snapshots were extracted from each case, yielding a training set size of 500. To study the effects of varying basis size, bases with $M = 4, \ldots, 10$ were constructed. For the studied scenario it was found that errors became impractically large for $M < 4$, and decreased extremely slowly for $M > 10$.

4.2 Performance of the Reduced Order Models

The full and reduced order models were subjected to three test load cases, randomly generated from the range of the training set loads.

Critical Time Step and Solution Time. The variation in the critical time step Δt_{cr} with basis size is plotted in Fig. 1. Generally, as the basis size M increases, the time step decreases, approaching the full model value. The concomitant improvement in solution time for each basis is shown also. For the current scenario, a peak speedup of 13.4 times was obtained for $M = 4$. Even using the largest basis ($M = 10$) resulted in a speedup of 4.3 times. The gap between the time step improvement and the corresponding solution time improvement is indicative of the greater computational cost *per step* of the new algorithm. We estimate the step-wise cost to be between 1.7 (for $M = 4$) and 2.3 (for $M = 10$) times greater for the new algorithm. As observed, though, this cost is outweighed by the step size increase nonetheless. The proportion of the overall computation time spent on each algorithm step is shown in Fig. 1, also.

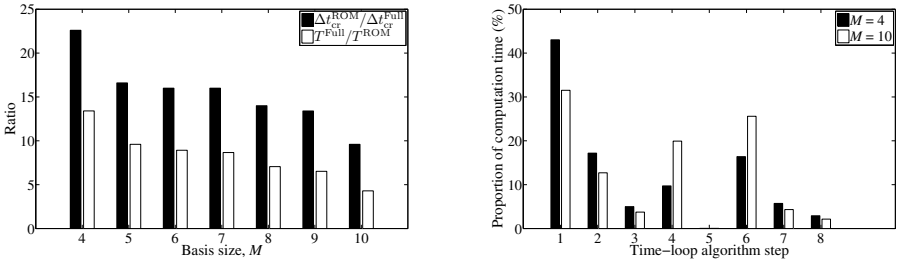


Fig. 1. (a) Left: Ratios of reduced order model critical time step $\Delta t_{\mathrm{cr}}^{\mathrm{ROM}}$ to full model step $\Delta t_{\mathrm{cr}}^{\mathrm{Full}}$, and ratios of full model overall solution time T^{Full} to reduced model time T^{ROM}, as functions of basis size. Right: Proportion of the computation time taken for each time-loop algorithm step (Sect. 3) for the smallest and largest bases.

Solution Errors. The nodal solution errors for each load case and each basis size are shown in Fig. 2. For $M < 7$, errors began to rise more sharply, though the

maximum value over all simulations was still < 0.6mm. Using $M \geq 5$, ensured relatively low errors in all cases, while still affording a solution speedup of 9.6 times. Given typical brain MR image resolutions are around 1mm, we feel these error levels are acceptable for medical imaging applications, and are, moreover, within the realistically obtainable accuracy of full models [3,7,10,11,14]. The distribution of error over the mesh for load case 3 is shown in Fig. (2), also.

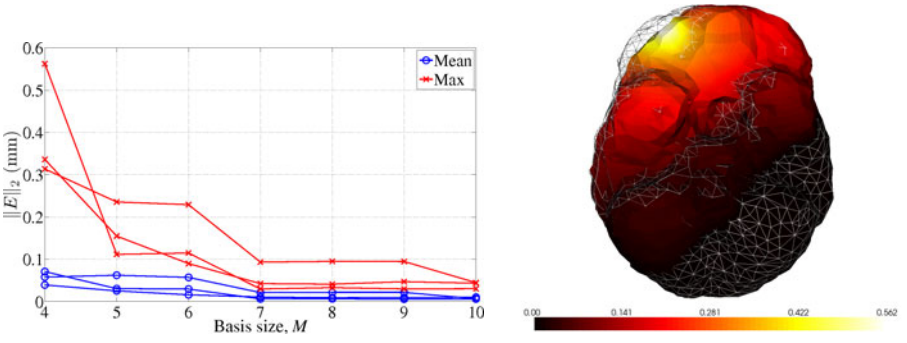


Fig. 2. Left: Mean and maximum nodal error magnitudes as a function of basis size for the three test load cases. Right: Inferior view of the deformed mesh for load case 3 (worst case) using the smallest basis ($M = 4$), with colours mapped to error magnitude (units are mm). A wireframe model of the undeformed mesh is overlaid.

5 Discussion and Conclusions

We have presented a reduced order nonlinear explicit finite element algorithm for soft tissue simulation, and described its implementation for graphics hardware execution. Using a basis size of $M = 5$, the method affords solution speedups of nearly an order of magnitude, while confining errors with respect to the full model to less than 0.24mm (corresponding to about 1-2% of the maximum deformation). For the example brain model investigated, overall solution times were approximately 500ms for a simulated time of 1sec – hence, real-time solution was comfortably achieved. While GPU-based execution was employed, it should be noted that these speed improvements result entirely from algorithmic modifications, not hardware differences.

A drawback of the current algorithm is its support for *homogeneous* essential boundary conditions, only. In this study, loading was applied via natural (force-based) boundary conditions. However, as noted in [6], methods exist for converting essential boundary conditions into equivalent natural ones. Penalty methods, which form the basis of widely used penalty contact formulations, are one such example. Our current work involves development of efficient procedures for this purpose.

Our approach may have applications in motion compensation for image-guided therapies, for which fast and accurate solutions are required. Particularly relevant examples include compensation for brain motion during neurosurgery,

as described, and for respiratory motion during lung radiotherapy [2]. In both applications, the loading is approximately known prior to the intervention, allowing reasonable training sets to be developed. Solutions for the precise loadings emerging intra-operatively may then be obtained rapidly and with a high level of confidence.

References

1. Carter, T.J., Sermesant, M., Cash, D.M., Barratt, D.C., Tanner, C., Hawkes, D.J.: Application of soft tissue modelling to image-guided surgery. Medical Engineering & Physics 27(10), 893–909 (2005)
2. Colgan, R., McClelland, J., McQuaid, D., Evans, P.M., Hawkes, D., Brock, J., Landau, D., Webb, S.: Planning lung radiotherapy using 4D CT data and a motion model. Physics in Medicine and Biology 53, 5815–5830 (2008)
3. Ferrant, M., Nabavi, A., Macq, B., Jolesz, F., Kikinis, R., Warfield, S.: Registration of 3-D intraoperative MR images of the brain using a finite-element biomechanical model. IEEE Transactions on Medical Imaging 20(12), 1384–1397 (2001)
4. Hill, D.L.G., Maurer, C.R., Maciunas, R.J., Barwise, J.A., Fitzpatrick, J.M., Wang, M.Y.: Measurement of intraoperative brain surface deformation under a craniotomy. Neurosurgery 43(3), 514–526 (1998)
5. Joldes, G.R., Wittek, A., Miller, K.: Non-locking tetrahedral finite element for surgical simulation. Communications in Numerical Methods in Engineering 25(7), 827–836 (2009)
6. Krysl, P., Lall, S., Marsden, J.E.: Dimensional model reduction in non-linear finite element dynamics of solids and structures. International Journal for Numerical Methods in Engineering 51, 479–504 (2001)
7. Miga, M.I., Paulsen, K.D., Lemery, J.M., Eisner, S.D., Hartov, A., Kennedy, F.E., Roberts, D.W.: Model-updated image guidance: initial clinical experiences with gravity-induced brain deformation. IEEE Transactions on Medical Imaging 18(10), 866–874 (1999)
8. Niroomandi, S., Alfaro, I., Cueto, E., Chinesta, F.: Real-time deformable models of non-linear tissues by model reduction techniques. Computer Methods and Programs in Biomedicine 91, 223–231 (2008)
9. NVIDIA Corporation: NVIDIA CUDA Programming Guide Version 2.3 (2009)
10. Skrinjar, O., Nabavi, A., Duncan, J.: Model-driven brain shift compensation. Medical Image Analysis 6(4), 361–373 (2002)
11. Tanner, C., Schnabel, J.A., Leach, M.O., Hose, D.R., Hill, D.L.G., Hawkes, D.J.: Factors influencing the accuracy of biomechanical breast models. Medical Physics 33(6), 1758–1769 (2006)
12. Taylor, Z.A., Cheng, M., Ourselin, S.: High-speed nonlinear finite element analysis for surgical simulation using graphics processing units. IEEE Transactions on Medical Imaging 27(5), 650–663 (2008)
13. Taylor, Z.A., Comas, O., Cheng, M., Passenger, J., Hawkes, D.J., Atkinson, D., Ourselin, S.: On modelling of anisotropic viscoelasticity for soft tissue simulation: numerical solution and GPU execution. Medical Image Analysis 13(2), 234–244 (2009)
14. Wittek, A., Miller, K., Kikinis, R., Warfield, S.K.: Patient-specific model of brain deformation: Application to medical image registration. Journal of Biomechanics 40(4), 919–929 (2007)

Simulation of Nodules and Diffuse Infiltrates in Chest Radiographs Using CT Templates

G.J.S. Litjens[1,2], L. Hogeweg[2], A.M.R. Schilham[2], P.A. de Jong[2], M.A. Viergever[2], and B. van Ginneken[2,1]

[1] Diagnostic Image Analysis Group, Radboud University Nijmegen Medical Centre, The Netherlands
[2] Image Sciences Institute, Utrecht University Medical Center, The Netherlands

Abstract. A method is proposed to simulate nodules and diffuse infiltrates in chest radiographs. This allows creation of large annotated databases for training of both radiologists and computer aided diagnosis systems. Realistic nodules and diffuse infiltrates were generated from three-dimensional templates segmented from CT data. These templates are rescaled, rotated, projected and superimposed on a radiograph. This method was compared, in an observer study, to a previously published method that simulates pulmonary nodules as perfectly spherical objects. Results show that it is hard for human observers to distinguish real and simulated nodules when using templates (AUC-values do not significantly differ from .5, $p > .05$ for all observers). The method that produced spherical nodules performed slightly worse (AUC of one observer differs significantly from .5, $p = .011$). Simulation of diffuse infiltrates is challenging but also feasible (AUC=0.67 for one observer).

Keywords: CT, radiograph, simulation, nodules, diffuse infiltrates.

1 Introduction

Lung diseases are among the largest causes of death and disability in the world. In most cases early detection, for example in screening using computed tomography (CT) or radiography, increases the chance of recovery significantly. Due to for example dose and cost considerations, chest radiography is by far the most common radiological exam [1]. It has been shown previously that in chest radiography detection of pulmonary nodules and the detection of focal shadowing is extremely difficult, even for expert radiologists [2]. For nodules it was found that when a nodule was missed, it was visible in retrospect in 90% of cases [3].

This shows how important it is to train radiologists in reading chest radiographs (CRs). Also, the development of computer aided diagnosis (CAD) systems to act as a first or second reader has been shown to improve accuracy [4]. For both radiologist and CAD training extensive databases with proven radiological findings are required, but constructing such databases is a laborious process.

In this paper a method is proposed to simulate lesions on CRs. This method uses CT templates of real nodules and diffuse infiltrates in combination with

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 396–403, 2010.

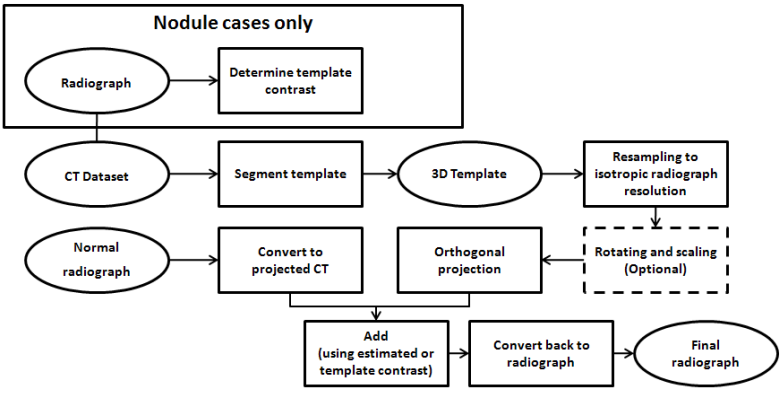


Fig. 1. Flowchart of the method for simulating abnormalities in radiographs

normal CRs which enables generation of large amounts of training data. Section 2 describes the data and Section 3 the method. In Section 4 the results of an observer study are presented that investigates if simulated lesions are indistinguishable from real pathology. Section 5 discusses the results and suggests topics for further research.

2 Materials

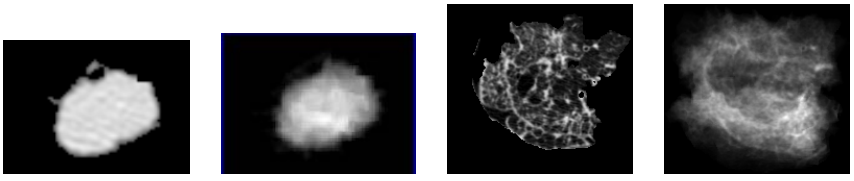
For the nodule templates radiographs and CT images from the NELSON lung cancer screening trial [5] were used. We used 36 pairs of radiographs and CT scans obtained within a three month period. The nodules in the radiographs were subsequently annotated by a radiologist using the CT scan as ground truth.

The templates for generating diffuse infiltrates were obtained from 44 clinical CT scans with corresponding radiographs. These radiographs were also annotated according to the findings in the CT scan. The annotation was done by an expert chest radiologist. [6]

Finally, 42 normal radiographs images were selected from the lung cancer screening database. All radiographs in this study were acquired using a Philips Digital Diagnost with a cesium iodine scintillator, a 3000 × 3000 pixels acquisition matrix and .143 mm resolution.

3 Methods

The proposed method is an extension and adaption of a previously published method by Schilham et al. [7]. The simulation method can be summarized (Figure 1) as follows: a radiograph and a lesion template (nodule or diffuse infiltrate segmented from a CT scan) are needed. The segmented lesion is resampled with 3D cubic interpolation to an isotropic resolution equal to the pixel resolution of



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

Fig. 2. Simulation templates: (a) CT slice of a segmented nodule, (b) Projection of a CT nodule, (c) CT slice of a segmented diffuse infiltrate and (d) Projection of a CT diffuse infiltrate

the radiograph. As an additional step the lesion can be rotated and scaled. Subsequently the pixel values of the radiograph are converted to projected Hounsfield units. The lesion is then projected to a 2D image and superimposed on a random location in this image using a contrast value. The final step is conversion from the projected CT values to pixel values of the radiograph.

CT template acquisition. The segmentation of the nodules from the CT was performed using a segmentation technique published by Kostis et al. [8]. In some cases a segmentation was deemed unusable because of the shape of the nodule, e.g. in the case of a pleural lesion. This resulted in a final number of 20 templates.

Diffuse infiltrates were segmented using a previously developed semi-automatic method. Twenty-six diffuse infiltrate templates were obtained. Examples of templates are shown in Figure 2.

Preprocessing and projection of the template. The obtained templates were resampled to an (almost) isotropic resolution equal to the resolution of the radiograph pixels using cubic 3D interpolation. Next the template can be rotated and/or rescaled to allow the creation of a wide range of different lesions from a single template.

The template has to be projected to a 2D image to allow superimposition of the template on a radiograph. Using simple raycasting the lesion template can be projected to a 2D image.

Conversion from radiograph to projected CT. Projection of a CT data set should in theory be similar to a radiograph. In practice there are several reasons why this is not the case, such as different x-ray energies and proprietary post-processing algorithms applied by manufacturers to radiographs. To overcome these problems, a conversion step from the pixel values encountered in the radiograph to projected CT is needed.

In this study a simple intensity mapping is used:

$$I_{\mathrm{CT}} = F\left[I_r\right], \tag{1}$$

where I_r and I_{CT} represent the intensity of a pixel in the radiograph and the projected CT, respectively, and F is the mapping function. This mapping

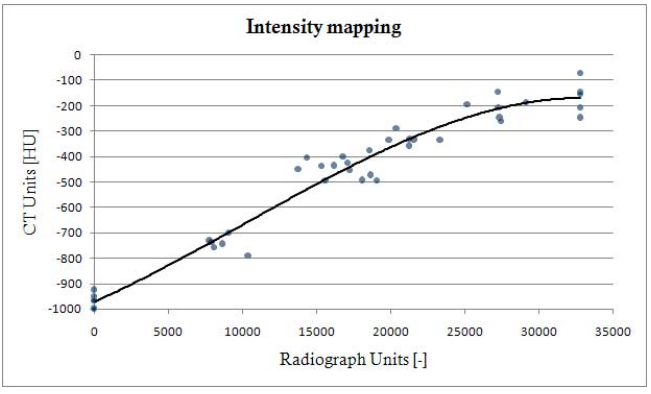


Fig. 3. Relation between intensities in radiographs and average projected CTs determined by a third-order polynomial. Data points were obtained from measurements in five radiographs and projected CT images.

function can be determined by measuring the intensities in both images in similar regions. The intensities were measured in the heart, in the spine and in the lung fields, and, additionally, the lowest and highest pixel value were added. These five values, determined in five radiographs and five corresponding projected CTs from the lung screening trial database, were used to fit a third-order polynomial, and this resulted in the mapping function, shown in Figure 3. It will not completely undo the effects of the more advanced post-processing steps (edge detection, unsharp masking) but it was adequate for the purpose of this study.

Superimposition. The final step is to add the projected template to the converted radiograph. Lung masks generated using a published method [9] were used to make sure that generated nodules were contained completely within the lung fields.

For diffuse infiltrate templates a different approach was used. A random position was selected within the lung mask. All the template pixels outside the mask were discarded. To remove edge artifacts, pixels close to the lung border (10 pixels, Euclidean distance) were added with reduced intensity (using an exponential falloff). The superimposition itself is straightforward. First a contrast measure is defined:

$$c = \log\left[\frac{I_t}{I_b}\right], \tag{2}$$

where c is the template contrast and I_t and I_b are the mean template and background intensities, respectively. Next, for each template the mean background intensity I_b and mean template intensity I_t were measured in the radiograph belonging to the CT scan the template was segmented from. This resulted in a contrast value c for each template. When superimposing the template on a radiograph, this c is used in conjunction with the mean background intensity at

the selected template position to determine what the mean template intensity I_t should be to achieve the same contrast. The template mean is then adjusted to this value. This allows realistic superimposition of templates.

For nodules it is easy to implement this automatically because the standard deviation on I_b is small due to the relatively small size of the templates. However, for diffuse infiltrate templates this is more difficult as they can span a large region of the lung, and the contrast can vary considerably within that region. Therefore the contrast values of simulated diffuse lesions were adjusted manually in this study. Some simulation results are shown in Figure 4.

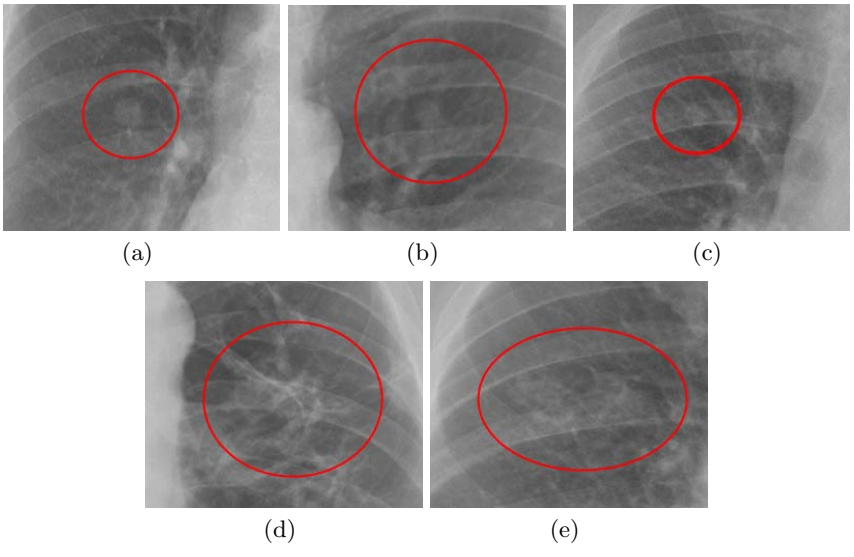


Fig. 4. Real lesions and simulation examples: (a) Real nodule, (b) Template based lesion simulation, (c) Perfectly spherical object nodule simulation, (d) Real diffuse infiltrate and (e) Simulated diffuse infiltrate

Perfect spherical nodule simulation. Samei et al. [10] have validated a different method for simulation of nodules in CRs. This method uses a mathematical representation of a nodule contrast profile. This profile is based on measurements in a real nodule database and the formula for this mask is:

$$c[r] = C(\frac{4}{D^4}r^4 - \frac{4.2}{D^2}r^2 + 1) \tag{3}$$

Here r is the position along the radius, C is the peak contrast value and D is the diameter. The definition of contrast is the same as in Eq. 2. Using a value for c and D, a simulated nodule can be added to a radiograph (Fig. 4c).

Table 1. Area under the curve and corresponding p-values (AUC / p-value) when compared to AUC=.5 for all observers

Observer	All nodules	Template method	Samei's method
1	.59 / .2	.49 / 1	.66 / .01
2	.40 / .1	.39 / .2	.41 / .2
3	.57 / .3	.56 / .5	.59 / .2

4 Experiments and Results

To investigate if the template based lesion simulation can generate realistic nodules and diffuse infiltrates in radiographs, two observer studies were performed. Three observers participated in the studies. Observer 1 and 2 were image analysis scientists with a expertise in chest CT and radiography. Observer 3 was a radiologist specializing in chest radiology. Whether a simulated lesion was realistic was determined using an analog scale from 1 to 100, represented as a slider to the observers.

In the first experiment 39 real nodules were used. Using template based lesion simulation a set of 39 nodules was simulated. In addition, using the perfectly spherical nodule simulation method, another 39 nodules were simulated. The parameters used for perfectly spherical nodule generation were a nodule diameter D between 7 and 22 mm and a peak contrast value C of between .15 and .22, which was comparable to our templates. This resulted in a total of 117 nodules, of which 78 are simulated, 39 for each method. In the second experiment the hypothesis that simulated and real diffuse infiltrates were indistinguishable is tested. To this end, 23 images containing diffuse infiltrates were acquired from clinical practice for which annotations by an expert radiologist were available. In these images between two and four regions were selected for scoring, leading to a total of 66 regions. In addition, diffuse infiltrate templates were superimposed on 23 normal radiographs, varying between 3 and 8 templates per image to get a good coverage of the lung. In these radiographs, 2 to 4 regions were annotated, leading to a total of 63 regions. The results are presented as receiver operating characteristic (ROC) curves in figure 5. A non-parametric method [11] is used to determine if the area under the curve (AUC) differs significantly from .5. Table 1 shows the AUC and p-values of all curves on nodule simulation.

Observer 1 was the only observer to achieve an AUC significantly larger than chance performance in the nodule experiments, when comparing real nodules with the nodules simulated as perfect spheres. This observer performed the study on diffuse infiltrates and attained an AUC of .67, significantly higher than .5 ($p = .001$), this is shown in figure 5d.

5 Discussion and Conclusion

The results show that it is difficult for the observers to distinguish real and simulated nodules. Although both observer 1 and 3 do have a AUC higher than .5

for the all-nodule case, the p-values indicate that this difference is not significant ($p > .05$). Figure 5b and c show the results for the methods separately. It can be seen that observer 1 and 3 perform better on nodules simulated with the perfectly spherical object nodule method. These nodules seem less realistic than those obtained by the template based lesion simulation method presented here. This can also be seen in table 1, where for the spherical nodules the AUC is significantly different from .5 ($p < .05$) for observer 1. No significant difference is found between the AUC and an AUC of .5 for nodules simulated with the template method for all observers ($p > .05$). Observer 1 was asked to participate in the diffuse infiltrate experiment as he performed the best on the nodule cases. His results show a significant difference between his AUC of .67 and an AUC of .5 ($p = .001$). This means that observer 1 can see the difference between some

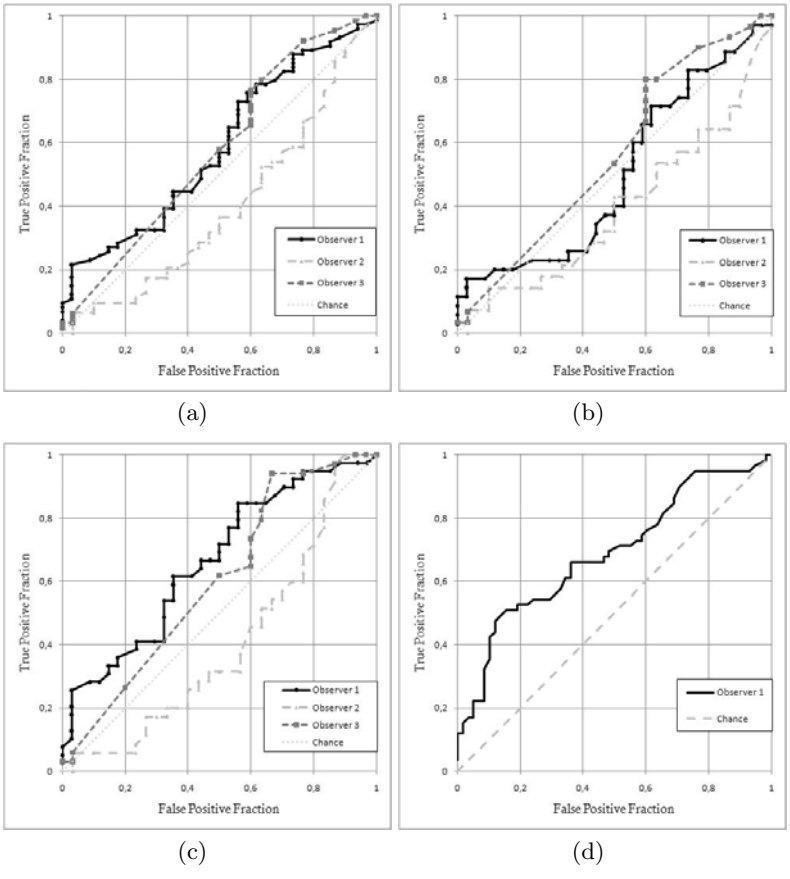


Fig. 5. ROC curves of the observer studies performed on distinguishing simulated lesions from real lesions: (a) Observer ROC curve of all nodules, (b) ROC curve of nodules simulated with the template method, (c) ROC curve of nodules simulated with perfectly spherical object method and (d) ROC curves for diffuse infiltrates

real and simulated diffuse infiltrates. As the AUC is still low, it seems plausible that diffuse infiltrates can be generated.

Future work should focus on generating a contrast function to allow better superimposition of especially diffuse infiltrates. Creating a realistic contrast between lesion and background is the most challenging aspect in these simulations. Also, the effect of post-processing could be investigated by using raw radiographs. It could also be helpful to collect more templates for diffuse infiltrates which exhibit a large variation in appearance.

Concluding, it is possible to simulate nodules in radiographs that are indistinguishable from real nodules. Using a small set of templates, a wide range of lesions can be created. For diffuse infiltrates, results show that many simulated lesions are realistic, which is an encouraging result. The proposed method is the first method that allows the simulation of diffuse infiltrates in CRs, and for simulating nodules it is more effective than previously published methods.

References

1. Bhargavan, M.: Trends in the utilization of medical procedures that use ionizing radiation. Medical Physics 134, 612–627 (2008)
2. Forrest, J.V., Friedman, P.J.: Radiologic errors in patients with lung cancer. Western Journal of Medicine 134, 485–490 (1981)
3. Muhm, J.R., Miller, W.E., Fontana, R.S., Sanderson, D.R., Uhlenhopp, M.A.: Lung cancer detected during a screening program using four-month chest radiographs. Radiology 148, 609–615 (1983)
4. Abe, H., Macmahon, H., Engelmann, R., et al.: Computer-aided Diagnosis in Chest Radiography: Results of Large-Scale Observer Tests at the 1996–2001 RSNA Scientific Assemblies. Radiographics 23, 255–265 (2003)
5. 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)
6. Arzhaeva, Y., Prokop, M., Tax, D.M.J., de Jong, P.A., Schaefer-Prokop, C.M., van Ginneken, B.: Computer-aided detection of interstitial abnormalities in chest radiographs using a reference standard based on computed tomography. Medical Physics 34(12), 4798–4809 (2007)
7. Schilham, A.M.R., van Ginneken, B.: Simulating nodules in chest radiographs with real nodules from multi-slice CT images. In: Medical Imaging (2006)
8. Kostis, W., Reeves, A., Yankelevitz, D., Henschke, C.: Three-dimensional segmentation and growth rate estimation of small pulmonary nodules in helical CT images. Trans. Med. Imag. 22(10), 1259–1274 (2003)
9. van Ginneken, B., Stegmann, M.B., Loog, M.: Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database. Medical Image Analysis 10(1), 19–40 (2006)
10. Samei, E., Flynn, M.J., Eyler, W.R.: Detection of Subtle Lung Nodules: Relative Influence of Quantum and Anatomic Noise on Chest Radiographs. Radiology 213(3), 727–734 (1999)
11. DeLong, E., DeLong, D., Clarke-Pearson, D.: Comparison of the area under two or more correlated receiver operating characteristic curves. Biometrics 44, 837–845 (1988)

High-Fidelity Meshes from Tissue Samples for Diffusion MRI Simulations

Eleftheria Panagiotaki[1], Matt G. Hall[1], Hui Zhang[1], Bernard Siow[1,2], 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 paper presents a method for constructing detailed geometric models of tissue microstructure for synthesizing realistic diffusion MRI data. We construct three-dimensional mesh models from confocal microscopy image stacks using the marching cubes algorithm. Random-walk simulations within the resulting meshes provide synthetic diffusion MRI measurements. Experiments optimise simulation parameters and complexity of the meshes to achieve accuracy and reproducibility while minimizing computation time. Finally we assess the quality of the synthesized data from the mesh models by comparison with scanner data as well as synthetic data from simple geometric models and simplified meshes that vary only in two dimensions. The results support the extra complexity of the three-dimensional mesh compared to simpler models although sensitivity to the mesh resolution is quite robust.

1 Introduction

Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) is a non-invasive imaging technique sensitive to the dispersion of water molecules in biological tissues. It has become an essential tool for probing microstructure. Diffusion Tensor Imaging (DTI) indices such as mean diffusivity and diffusion anisotropy provide useful but non-specific markers of white matter integrity. More recent techniques [1,2,3,4] aim to estimate specific features of tissue microstructure such as axon diameter and density. However, questions remain about how well these microstructural indices from diffusion MRI reflect the actual tissue microstructure since it is difficult to obtain ground truth information.

Synthetic data is a powerful tool for developing, optimizing, evaluating and comparing diffusion MRI techniques. Unlike scanner data, it provides a ground truth, thereby allowing performance evaluation of methods in a controlled environment. One approach for generating synthetic data is to use a simple model like the DT [5] which describes the displacement of particles with a Gaussian distribution. This simple model provides measurements with negligible computation cost. However, due to its simplicity the DT model ignores features of the tissue and the diffusion process within. In particular, it does not account

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 404–411, 2010.

for restricted diffusion within cells so oversimplifies the signal from real tissue. To address this limitation, recent diffusion MRI studies [1,2,3,4,6,7] use a variety of multi-compartment models that separate restricted, hindered and free diffusion. Analytic compartmental models are computationally efficient but are approximations and limited to describing diffusion in simple geometries such as cylinders and spheres.

This has motivated work on numerical methods, which allow arbitrary complexity in the diffusion environment and measurement process. The most common numerical methods for synthesizing diffusion MRI data are based on finite-difference approaches e.g. [8] and Monte-Carlo methods e.g. [9,10,11]. Finite-difference methods rely on an approximate solution to the diffusion equation at a discrete number of points, e.g. on a grid, while Monte-Carlo methods simulate Brownian motion of spins within geometric models of tissue. Numerical methods can investigate diffusion environments that are impossible using analytic models. For example, Hall and Alexander [10] use Monte-Carlo simulations to synthesize data in a model of tissue undergoing swelling to simulate oedema where cylindrical axons swell and abut.

Other work [8,11] base the simulations in models derived from light microscopy images for a better approximation of the tissue. Lipinski [11] was the first to use histologic images in combination with two-dimensional Monte-Carlo simulations to study the diffusion signal. The tissue model is based on a rough segmentation of a digitized light microscopy image of white matter tissue. More recently, Chin et al. [8] use a finite-difference approach and construct a tissue model by replicating a light microscopy image in the third dimension. These numerical models are an improvement on the simplified geometric models, yet, the single-slice microscopy images are low-resolution and the models describe the tissue only in two dimensions. However, numerical methods are more computationally intensive than analytic models and the implementation of complex realistic environments is non-trivial, causing optimization and tuning of the simulation to be especially challenging.

This paper presents a method for constructing three-dimensional tissue models from a stack of high-resolution Confocal Laser Scanning Microscopy (CLSM) images. We capture the three-dimensional structure of biological tissue in a more natural way than previous studies [8,11]. We demonstrate the method using a biological phantom (asparagus) which is a useful model with similar microstructure to white matter [12]. Experiments vary simulation parameters and mesh properties to optimize the precision of the synthesized data while minimizing computational complexity. We compare results to scanner data and to synthetic data from simple parametric models and two-dimensional meshes.

Section 2 describes the method for constructing and evaluating three-dimensional tissue models. Section 3 outlines the experiments for optimizing simulation parameters, exploring the effect of mesh resolution, and evaluating synthetic data from the mesh by comparison with scanner data and with data generated from simpler models. Section 4 concludes.

2 Methods

The full procedure has several stages. First, we acquire DW-MRI data with a wide range of diffusion times and diffusion weightings with gradient directions both parallel and perpendicular to the asparagus stem. We identify a region of interest (ROI) in the DW-MRI data containing one of the vascular bundles (Fig.1a), which we cut from the stem and image with CLSM (Fig.1b) to obtain a stack of images. The vascular bundles in the asparagus consist of highly-organised cylindrical fibres with thick walls that exhibit anisotropic diffusion and the distribution of capillary sizes is similar to brain white matter tissue [12]. We construct the three-dimensional mesh model (Fig.1c) with the marching cubes algorithm [13] and use it as a substrate in Monte-Carlo simulations [10] to synthesize DW-MRI data.

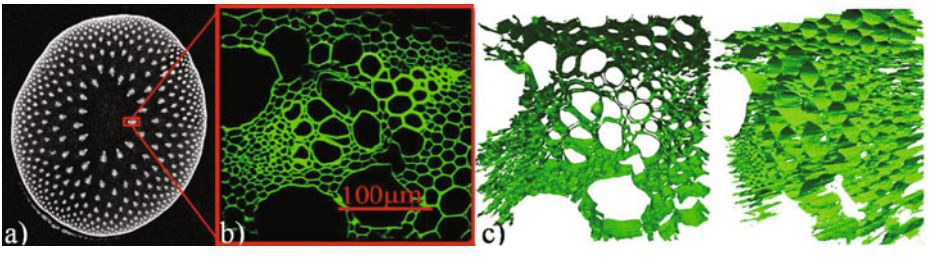


Fig. 1. a) DW-MRI image of a transverse section of the asparagus stem. The red square indicates the ROI, containing one of the vascular bundles, appearing white in the MRI image. **b**) CLSM image of the same ROI. **c**) three-dimensional mesh model shown from two different angles.

Sample preparation: We place a stem of green asparagus (*Asparagus officinalis*) in a syringe padded with cotton soaked in pure water. This keeps the sample hydrated, preventing shrinkage and diffusion changes during the scan.

MRI Acquisition: We acquire DW-MR images with a small bore 9.4T scanner (Varian) with maximum gradient strength 400 mT/m and use a controlled air-flow mechanism to keep the sample temperature constant within $\pm 1\,^{\circ}$C. The two-direction encoding scheme has one direction parallel to the asparagus stem and one perpendicular. We acquire 64 pulsed-gradient spin-echo (PGSE) measurements with six diffusion times, $\Delta = 10, 30, 50, 70, 80, 100$ ms, three gradient durations $\delta = 3, 10, 20$ ms and gradient strength $|\mathbf{G}|$ varied from 40 to 400 mT/m in ten steps of 40 mT/m. The 64 measurements includes all combinations with $\Delta > \delta$ and $b < 6.5 \times 10^9\,\mathrm{sm}^{-2}$. We use the minimum echo time (TE) possible for each measurement and set the repetition time (TR) to 3 s. The total acquisition time is approximately 40 hours. We correct for T_2 dependence by acquiring separate $b = 0$ images for each combination of δ and Δ. The in-plane field of view is 16 mm. The matrix size is 256×256 and the slice thickness is 0.5 mm.

Confocal Acquisition: We cut out four 600 μm thick sections which we then stain with Eosin for 10 minutes and wash thoroughly with phosphate buffered saline. To avoid any geometric distortions we use vibratome, which is designed for sectioning soft tissue without freezing or paraffin, and choose to image only the middle slices. We use a Leica SP2 AOBS confocal multi-photon laser scanning microscope coupled to a Leica DMRE upright microscope (Leica, Milton Keynes, UK). We receive the laser output with an electro-optical modulator (EOM) (Linos LIV20) before delivering to the confocal microscope through a series of optical mirrors. The EOM allows the laser intensity at the objective to be controlled and optimized. The EOM is set at 90% for imaging to ensure that the polarization of the incidental laser beam remained consistent across all specimens. We image the specimens with a 40×1.25NA oil Plan Apo objective to give image dimensions of $315\,\mu$m $\times$ $315\,\mu$m. We acquire optical z-sections of $1\,\mu$m thickness reaching a maximum depth of $100\,\mu$m with an image averaging set to 3 per z-slice. The image size is 1024×1024 pixels.

Model Construction: To construct the mesh models we assemble the images into a stack and segment them by thresholding to create binary images that separate the intra- and extra-capillary space. The intra-cellular volume fraction is determined by counting the pixels below the threshold. We use the marching cubes algorithm [13] on the binary stack to produce the three-dimensional mesh model. Memory limitations require that we downsample the 100 CLSM images to 144×144 pixels while keeping the three-dimensional aspect ratio of the voxels the same as the original image stack, which also makes computation times manageable. The algorithm produces a mesh of around 500,000 triangles.

Simulations: The simulation system in [10] generates synthetic measurements for each combination of scan parameters from diffusing spins constrained by the structure of the mesh. Each triangle in the mesh acts as an impermeable reflecting boundary. For computational efficiency the system checks only for intersections with triangles within the range of each step. Spins are initialized uniformly across the mesh, in both intra- and extra-cellular regions. Here we assume the same properties (i.e. diffusivity, relaxation times) for both regions.

3 Experiments

This section outlines three sets of experiments. The first identifies a suitable combination of number of spins and updates within a fixed simulation runtime. The second explores the effect of the mesh resolution. The final experiment compares synthetic data from the three-dimensional mesh with scanner data and synthetic data from simpler geometric models.

3.1 Simulation Parameter Optimization

The aim is to maximize precision and accuracy of the synthetic measurements while minimizing computational cost. The complexity of the simulation is order $U = NT$ where N is the number of spins and T is the number of updates.

Without a ground truth, the accuracy of the simulation is hard to establish. However, [10] shows on simpler substrates that with fixed U accuracy tends to increase up to a certain N and suddenly depart although standard deviation decreases. Here we search for the same trend to identify the best trade off.

We use the 144×144 mesh as substrate with the same acquisition protocol as the scanner data and diffusivity $d = 2.1 \times 10^{-9}\,\mathrm{m}^2/\mathrm{s}$, estimated from scanner measurements with low b value using the monoexponential model. Simulations with various combinations of N, T for $U = 10^8$ are repeated 10 times with different initial spin positions. Each simulation runs in approximately 48 hours.

Results: The mean signal for the perpendicular direction over all scan parameter combinations and repeats remains similar as N increases from low $N \simeq 1 \times 10^4$ until $N \simeq 1 \times 10^5$ for $b = 1.5 \times 10^9\,\mathrm{sm}^{-2}$. All the measurements with different b values have similar trends with small variations that do not affect the final choice.

Conclusions: As in [10], we observe a gradual increase in reproducibility as N increases. The mean signal remains consistent from low N until $N \simeq 9 \times 10^4$ above which it increases noticeably. The increase is most likely a bias introduced by the timesteps being too long. We choose $N = 8 \times 10^4$ and $T = 1250$ to keep reproducibility high while remaining safely within the region of short enough timesteps.

3.2 Signal Dependency on Mesh Fidelity

This experiment compares meshes constructed from different resolution CLSM images to investigate the effect of varying mesh resolution on synthetic DW-MRI data. The highest image resolution we use is 144×144 pixels and the lowest 9×9 pixels with four intermediate stages. We synthesize data from the different resolution meshes using the parameters from experiment 3.1. We calculate the Mean Squared Error (MSE_{144}) compared to the 144×144 mesh to show differences in synthesizing data with different resolution meshes, and the MSE_{MRI} compared to the scanner data to reveal which of these differences are significant.

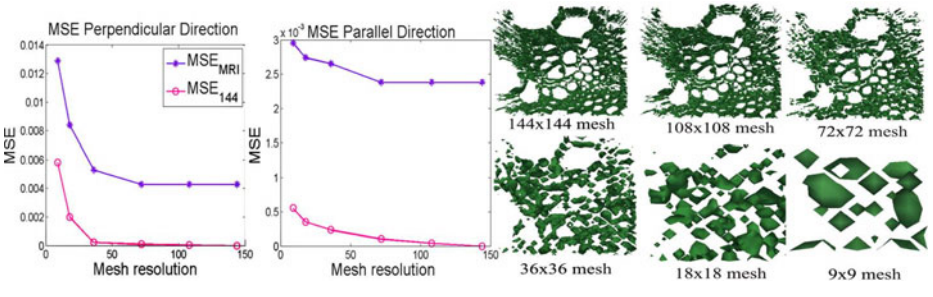


Fig. 2. Left: Plots of the MSE of the signal in the parallel and perpendicular direction in comparison with the high-resolution mesh and the scanner data against mesh resolution. Right: Illustration of the different complexity meshes used in the simulations.

Results: Figure 2 plots the MSE_{144} and MSE_{MRI} of the normalised signal for both directions against mesh resolution. We exclude all measurements with $S < 0.1$ from the plots and the MSE calculations e.g. to avoid significant noise-floor effects. In the perpendicular direction the MSE_{144} shows very little variation between the 144×144 and 36×36 resolution meshes. In the parallel direction, the MSE_{144} starts to increase at 72×72 resolution. The MSE_{MRI} however, shows little difference in both directions for the meshes with resolution 144×144 to 72×72.

Conclusions: In both directions the MSE_{MRI} remains similar for the meshes with resolution 144×144 to 72×72 meshes. At 36×36 we see slight elevation that becomes more obvious in lower resolution meshes. The results suggest that 72×72 resolution approximates the geometry closely enough to capture variations in water diffusion that MR signals are sensitive to.

3.3 Model Comparison with Scanner Data

The last experiment assesses the quality of synthetic MRI data of the mesh models. For comparison, we generate three sets of synthetic data: the 72×72 three-dimensional from experiment 3.2, an extruded two-dimensional mesh models and a packed-cylinder substrate with constant radius. To construct the extruded mesh we choose an image from the stack we used for the three-dimensional mesh model with $f = 0.8$ which is consistent with the three-dimensional model, replicate it to have the same number of slices used to generate the 72×72 mesh and run the same meshing algorithm. The packed-cylinder substrate has square-packed cylinders with radius $25\,\mu$m. The choice of single radius comes from the mean capillary radius in the mesh weighted by capillary volume [4]. We pick the packing density so that the intra-capillary volume fraction is also $f = 0.8$.

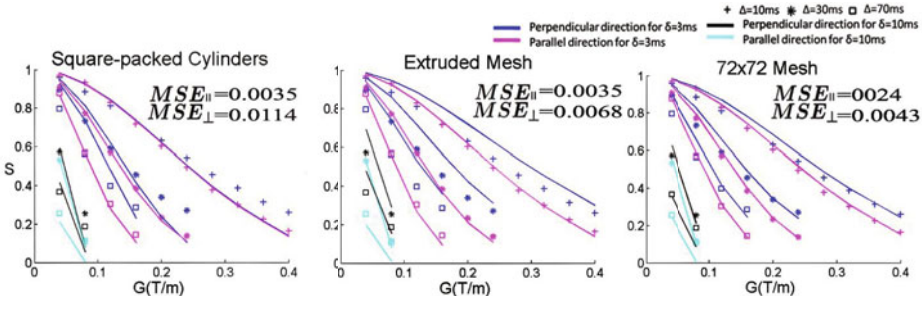


Fig. 3. Data synthesized from the cylinder model, the extruded mesh and the three-dimensional mesh with the scanner data from the PGSE experiment. For clarity, the normalised signal S is plotted only for selected values of Δ and δ as a function of the gradient strength $|\mathbf{G}|$ for the parallel and the perpendicular direction. Markers show the scanner data; lines show the synthetic data. Mean-squared error, for the parallel and perpendicular direction, are included for each model.

Results: Figure 3 compares data synthesized from the three models to the scanner data by plotting the normalised signal S only for selected values of Δ and δ as a function of the gradient strength $|\mathbf{G}|$ for the parallel and perpendicular direction. Again we exclude all measurements with $S < 0.1$ from the plots and the MSE calculations to avoid significant noise-floor effects. Predictions from simulations with the cylinders of constant radii and the extruded mesh model are unable to capture the data as well as the mesh model. For example the cylinders underestimate the perpendicular signals with $\Delta = 10,\ 30$ ms and $\delta = 3,\ 10$ ms while the extruded model overestimates them. The three-dimensional mesh model agrees closely in both directions for $\Delta = 10,\ 30,\ 70$ ms and $\delta = 3,\ 10$ ms. The MSE is lower for the three-dimensional mesh model in both directions.

Conclusions: The complex three-dimensional model minimizes the MSE and captures the data in both directions for $\Delta = 10,\ 30,\ 70$ ms and $\delta = 3,\ 10$ ms.

4 Conclusions

This work introduces a method for constructing a detailed tissue mesh model using CLSM to generate realistic diffusion MRI data. We investigate optimal simulation and mesh properties for precision and accuracy of the synthesized data. We test the simulated data from three-dimensional mesh models against scanner data, simpler extruded mesh models and simple parametric models. Results with the three-dimensional mesh model are very promising, agree with scanner data well and match the data better than the extruded and parametric models in both directions.

The method we propose in this paper can be refined in a number of ways. So far, we optimize the simulation parameters without a ground truth. In a similar experiment in [10] they compare results to an analytic model of restriction, however the mesh model here is much more complex and analytic solutions cannot provide a ground truth. Another aspect we could further explore is the choice of diffusivity. We could refine agreement by searching for the diffusivity that minimizes error between the synthetic data and the scanner data. The quality of the generated mesh also depends on the segmentation process and the downsampling of the confocal images. So far we use a simple thresholding algorithm. We could improve the accuracy and smoothness of the mesh by using more sophisticated segmentation algorithms. To avoid downsampling while preserving the topology with the minimum number of triangles we could use a meshing algorithm combined with decimation [14]. However, preliminary experiments with decimation reveal problems with intersecting triangles that cause the simulation to fail. Exploitation of more sophisticated meshing remains a focus for further work.

The Monte-Carlo simulation approach with a tissue model of high fidelity provides a mechanism for high quality data synthesis for testing and developing diffusion MR algorithms. It also allows for testing of the influence of subtle effects such as permeability and surface-particle interactions (e.g. particles temporarily trapped on the surface) that analytic models cannot capture. Future work will

extend the experiments to other samples in particular to brain tissue to allow testing and comparison of analytic models in a similar way to [7].

Acknowledgements. Thanks to Dr Chris Thrasivoulou for help with the confocal acquisition. Funded by EPSRC grants EP/E056938/1, EP/E007748, British Heart Foundation and BBSRC.

References

1. Stanisz, G.J., Szafer, A., Wright, G.A., Henkelman, R.M.: An analytical model of restricted diffusion in bovine optic nerve. Magnetic Resonance in Medicine 37, 103–111 (1997)
2. Assaf, Y., Basser, P.: Composite hindered and restricted model of diffusion (CHARMED) MR imaging of the human brain. NeuroImage 27, 48–58 (2005)
3. Alexander, D.C.: A general framework for experiment design in diffusion MRI and its application in measuring direct tissue-microstructure features. Magnetic Resonance in Medicine 60, 439–448 (2008)
4. Alexander, D.C., Hubbard, P.L., Hall, M.G., Moore, E.A., Ptito, M., Parker, G.J., Dyrby, T.B.: Orientationally invariant indices of axon diameter and density from diffusion MRI. NeuroImage (2010), doi:10.1016/j.neuroimage.2010.05.043
5. Basser, P.J., Mattiello, J., LeBihan, D.: MR diffusion tensor spectroscopy and imaging. Biophysical Journal 66, 259–267 (1994)
6. Behrens, T.E.J., Woolrich, M.W., Jenkinson, M., Johansen, H.: Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magnetic Resonance in Medicine 50, 1077–1088 (2003)
7. Panagiotaki, E., Fonteijn, H., Siow, B., Hall, M.G., Price, A., Lythgoe, M.F., Alexander, D.C.: Two-Compartment Models of the Diffusion MR Signal in Brain White Matter. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 329–336. Springer, Heidelberg (2009)
8. Chin, C.L., Wehrli, F.W., Hwang, S.N., Takahashi, M., Hackney, D.B.: Biexponential diffusion attenuation in the rat spinal cord: Computer simulations based on anatomic images of axonal architecture. Magnetic Resonance in Medicine 47, 455–460 (2002)
9. Szafer, A., Zhong, J., Gore, J.C.: Theoretical model for water diffusion in tissues. Magnetic Resonance in Medicine 33, 697–712 (1995)
10. Hall, M.G., Alexander, D.C.: Convergence and Parameter Choice for Monte-Carlo Simulations for Diffusion MRI. IEEE Transactions on Medical Imaging 28, 1354–1364 (2009)
11. Lipinski, H.G.: Monte Carlo simulation of extracellular diffusion in brain tissues. Phys. Med. Biol. 35, 441–447 (1990)
12. Lätt, J., Nilsson, M., Rydhög, A., Wirestam, R., Ståhlberg, F., Brockstedt, S.: Effects of restricted diffusion in a biological phantom: a q-space diffusion MRI study of asparagus stems at a 3T clinical scanner. Magnetic Resonance Materials in Physics, Biology and Medicine 20(4), 213–222 (2007)
13. Lorensen, W.E., Cline, H.E.: Marching cubes: A high resolution 3D surface construction algorithm. Computer Graphics 21(4) (1987)
14. Schroeder, W., Zarge, J., Lorensen, W., et al.: Decimation of triangle meshes. CG NY ACM 26, 65 (1992)

A Dynamic Skull Model for Simulation of Cerebral Cortex Folding

Hanbo Chen[1], Lei Guo[1], Jingxin Nie[1], Tuo Zhang[1], Xintao Hu[1], and Tianming Liu[2]

[1] School of Automation, Northwestern Polytechnical University, Xi'an, China
[2] Department of Computer Science and Bioimaging Research Center, The University of Georgia, Athens, GA, USA

Abstract. The mechanisms of human cerebral cortex folding and their interactions during brain development are largely unknown, partly due to the difficulties in biological experiments and data acquisition for the developing fetus brain. Computational modeling and simulation provide a novel approach to the understanding of cortex folding processes in normal or aberrant neurodevelopment. Based on our recently developed computational model of the cerebral cortex folding using neuronal growth model and mechanical skull constraint, this paper presents a computational dynamic model of the brain skull that regulates the cortical folding simulation. Our simulation results show that the dynamic skull model is more biologically realistic and significantly improves our cortical folding simulation results. This work provides further computational support to the hypothesis that skull is an important regulator of cortical folding.

Keywords: Cortex Folding, Simulation, Skull Constraint.

1 Introduction

The folding pattern of the cerebral cortex varies greatly between individuals [1], and it is believed to have certain relationship with the brain's cytoarchitecture and functional regionalization [2]. The factors that cause these differences have intrigued neuroscientists for a long period. Due to the difficulties in biological experiment and data acquisition for the developing fetus brain, there are growing efforts in the area of computational modeling and simulation that aim to understand the mechanisms of cortical folding. For example, Raghavan et al. proposed a continuum mechanics-based model of growth to synthesize cortical shapes by using physical laws [3]. Toro et al. proposed a computational morphogenetic model to study the fundamental mechanisms of cortical folding [4]. In our recent work [5], by performing a 3D morphogenetic model, we demonstrated that folding pattern is dependent on mechanical constraints of skull, cell growth rate, and initial geometry of the cortex [5]. We applied surface modeling to simulate the cerebral cortex morphogenesis and use a static skull model to constrain the growth of cerebral cortex.

In this paper, a dynamic skull model is developed to provide more biologically meaningful boundary condition for the cortical folding simulation. Mechanical constraint was the first major factor considered when investigating the determinants of cortical folding [6]. The hypothesis is that since the cortical area is almost three times

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 412–419, 2010.

larger than the cranial area, the cortex had to convolve to fit into a relatively small cranial volume. In previous skull models such as the ones in [3] and [5], the constraint force acts on the surface only when it touches the skull. While in this work, the skull constraint also considers the cerebrospinal fluid (CSF) pressure, as illustrated in Fig.1. Instead of using a cranial skull, we applied a dynamic growing volume to constrain the growth of cortex. Since the skull develops with the brain, it is more reasonable to model and simulate a dynamic growth skull constraint. The experimental results in this paper show that the dynamic skull model improves the cortical folding simulation significantly.

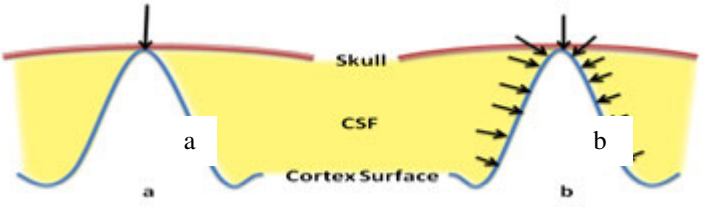


Fig. 1. (a) In previous skull models, constraint only acts on cortex surface when it touches the skull. (b) As the cortex is surrounded by CSF, the CSF pressure is also considered.

2 Method

The flowchart of our model and simulation is outlined in Fig.2. We reconstruct cortical surface from diffusion tensor imaging (DTI) data [7] and decompose it into multi-resolution representations via the spherical wavelet method [8]. The smooth cortical surface of the lowest resolution is used as the synthesized fetus cortex for the folding simulation. The folding simulation is driven by the neuronal growth model in [5] and regulated by the proposed dynamic skull model. The cortical shapes generated by the simulation models are evaluated by quantitative descriptors of curvature, parametric cortical folding measurement [9], and surface distance. It is noted that DTI data, instead of T1 structural data, is used here for surface reconstruction because this will facilitate our future investigation of how axonal fibers regulate the cortical folding.

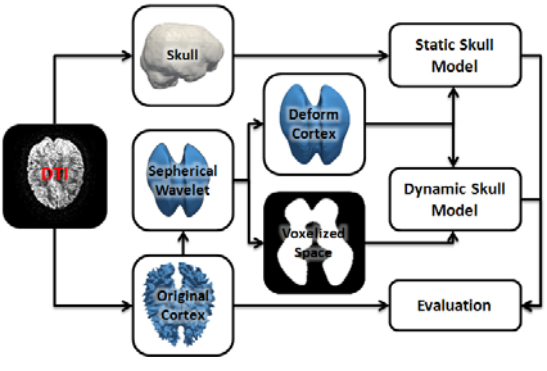


Fig. 2. The flowchart of the model and simulation

2.1 Materials and Pre-processing

Since the acquisition of fetus brain imaging data is quite challenging, we synthesize the fetus brain surface as follows. The original cortical surface was reconstructed from DTI data of adult brains using the DTI-based tissue segmentation [7] and the cortical surface reconstruction method in [10]. Then, we decompose the cortical surface into multi-resolution representations via the spherical wavelet algorithm [8],

in which the highly convoluted and complex cortical surface is then decomposed into a cascade of lower-resolution surfaces, as shown in Fig.3(a-e). Fig.3(f) shows a fetus cortical surface reconstructed from fetus MRI data [5]. We can see that the low-resolution surface (Fig.3(e)) appears to be similar to the fetus cortex in Fig.3(f). This visualization supports the correctness of our method for synthesizing fetus cortical surface. Moreover, in this way, we are able to compare the simulation results with original surface to evaluate the model.

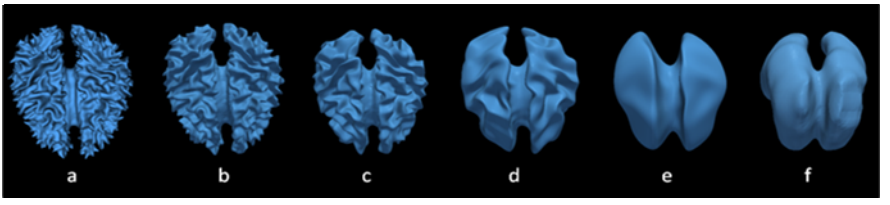


Fig. 3. (a) Original surface generated from DTI data. (b)-(e) Multi-resolution decomposition of cortical surface (the fifth, forth, third and second resolutions). (f) Cortical surface of fetus brain reconstructed from real MRI data [5].

Notably, the surface decomposition method [8] might produce unbalanced surface at low resolution, in which triangles in certain region might be much smaller than other regions as shown in Fig.4 or intersect with other triangles. Thus, to obtain a balanced initial surface, we voxelized the low-resolution surface into a volumetric image, and reconstructed a new surface using marching cube algorithm.

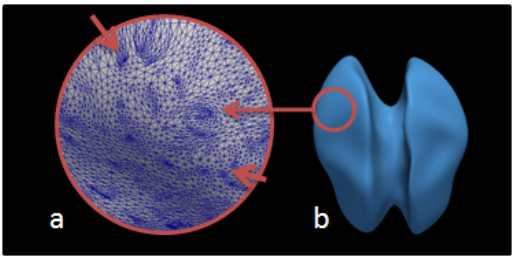


Fig. 4. The flowchart of deformable surface initialization. (a) Second resolution of the decomposed surface. (b) Voxelized surface. (c) Deformable surface. (d) Zoomed-in view of the surface patch in the circle in (a).

2.2 Static Skull

The development of the cerebral cortex is constrained by the cranial volume. We developed a volumetric constraint model [5] to simulate the effect of static skull during the folding of the cortex, as illustrated in Fig.5. When any vertex of the surface is being deformed to a new position $\mathbf{x}'$, the following condition should be satisfied: the new position of the vertex cannot intersect with

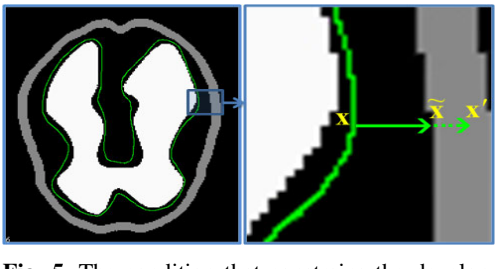


Fig. 5. The condition that constrains the development of cortex. When cortex point $\mathbf{x}$ is trying to deform to the new position $\mathbf{x}'$, the following condition should be satisfied: $\mathbf{x}'$ should not intersect with other brain tissues.

cranial volume. If the condition is not satisfied, another new position $\tilde{\mathbf{x}}$ for this vertex that satisfies the condition should be identified. After finding the new position satisfying the conditions, the vertex can be deformed to the new position.

2.3 Dynamic Skull

Since skull is developing during the cortical folding, its growth should be modeled dynamically. Thus, the development of skull that is caused by the outgrowth of cortex is considered in our paper, and the dynamic growth of skull is defined as follows. Similar to the volumetric constraint model in Fig.5, a constraint function $L(n)$ is defined on each voxel in the deformable space, where n is the iteration number. When $L(n) \le 0$, the voxel is defined as a deformable voxel, which cortex surface could deform into at iteration n, otherwise it is defined as constraint voxel. When $n = 0$, $L(n)$ is the initial skull constraint in which zero value space is defined as inner space of initial cortical surface in our experiments. When $n > 0$, $L(n)$ is interacting with the cortex folding process by:

$$L(n) = L(0) - \lambda_L \sum_n I(n) \tag{1}$$

where λ_L controls the growth speed of the skull and $I(n)$ is defined on each voxel. At iteration n, if any vertex or triangle on the cortex surface is trying to deform to a voxel, the value of $I(n)$ in this voxel is set as one, otherwise it is zero, meaning that the attempt of the cortex deformation into skull voxel will cause the growth of intracranial volume.

In this model, the constraint not only acts on the top of gyrus, but also on the lateral surface. The cortical surface is encircled by constraint voxel, thus when it grows to the outer space, its movement will be limited. This effect is the same as fluid pressure. As shown in Fig.1, because brains are surrounded by CSF, the constraint pressure should take the CSF pressure into consideration.

2.4 Development of the Cerebral Cortex

We used the models in [5] to simulate the development of the cerebral cortex. The developed model is composed of the following four key components.

a) *The deformable model.* The elasto-plasticity property is adopted on each edge of the triangle in the surface. Since the surface could be considered as a zero thickness sheet, the bending energy is introduced to model the rigidity of cortex.

b) *The growth model.* Growth of cerebral cortex is defined on each triangle of the surface. The classic logistic-growth function is adopted to describe the growth of cortical tissues as:

$$\frac{dA_{c0}}{dt} = A_{c0} m \left(1 - \frac{A_{c0}}{k}\right) \tag{2}$$

where m is known as the Malthusian parameter, k is the carrying capacity of the system and A_{c0} is the rest area of triangle.

c) *Constraint model*. The development of cortex is limited by the boundary condition of cranial volume.

d) *Model solver.* The proposed model can be formulated as a time varying partial differential equation. The explicit Newmark scheme is adopted to solve the model.

3 Experimental Results

In this section, we simulate the proposed computational model of cortical folding with static and dynamic skull constraints, and compare the results. Furthermore, we compare the simulations with different cortical growth rates.

3.1 Comparison of Two Skull Models

The values of parameters in the growth model are the same in the static skull model and dynamic skull model: λ_L =0.1, m =0.009, k =3. The results of folding development are illustrated in Fig.6. Fig.6(a-b) show the snapshots of cortex development at the iteration number 50, 100, 150, and 200 in static skull model and dynamic skull model, respectively. By visual evaluation, it is evident that more realistic folding is generated using the dynamic skull model. To quantify this difference, we use the average absolute Gaussian curvature and parametric folding pattern [9] as shape descriptor to evaluate the produced convolutions. We also use the distance between original surface and deformed surface to evaluate the simulation models. The differences between the average absolute Gaussian curvatures, folding pattern descriptors of the cortex surface in different simulation iterations, and the surface distances of five cases we simulated are illustrated in Fig.7(a-c) respectively. It is evident that the folding patterns produced by the dynamic skull models are much closer to those of normal brain surfaces obtained from real MRI images. It is also intriguing that when iteration number reaches certain steps, the average absolute Gaussian curvature of cortical surface will reach a steady condition. This steady condition is mainly because of the fact that after the cortex convoluted completely, the cortex will only develop vertically and thus curvatures do not change much. Fig.7(a) shows that the dynamic skull model needs only half of the iteration time of the static skull model to reach the same average curvature.

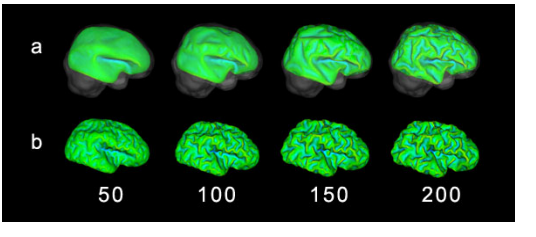


Fig. 6. Demonstrations of cortical development results: (a) Static skull model. (b) Dynamic skull model. (a)-(b) show the snapshots of cortex at the iteration number 50, 100, 150, and 200 respectively. A video of this progress is affiliated in supplementary materials.

3.2 Comparison of Growth Rates

In this section, we investigate the effect of skull growth rate on the cortical folding process using the dynamic skull model. Three groups of simulation were performed

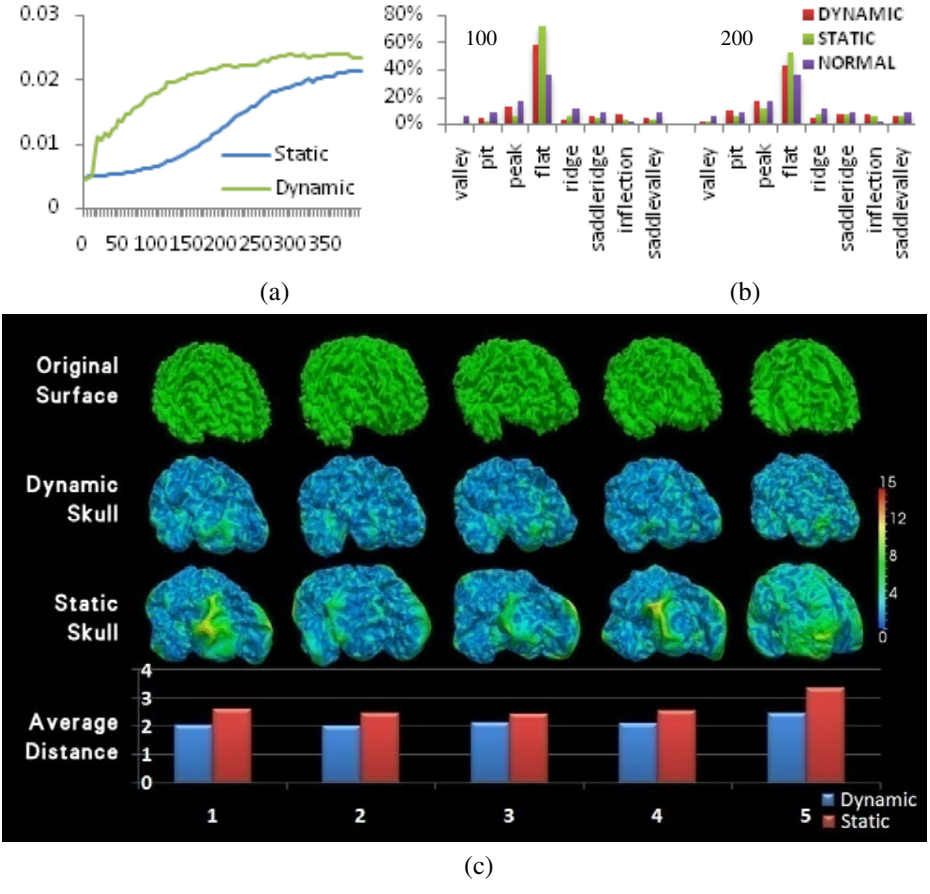
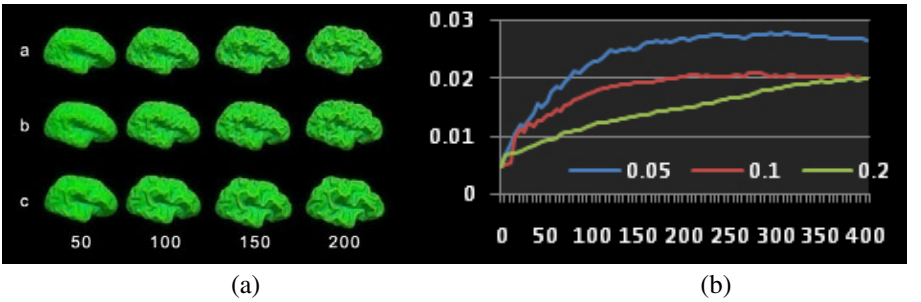


Fig. 7. (a) The difference of the average absolute Gaussian curvature of the cortex surface during simulated growth. (b) Folding pattern distributions at 100 and 200 iterations. More details of the folding descriptors are referred to [9]. (c) The distance between deformed surface and original surface for both dynamic and static skull models. The color bar is on the right.



(a) (b)

Fig. 8. (a) Cortical developing in different skull growth rates. a: λ_L =0.05; b: λ_L =0.1; c: λ_L =0.2. (b) The difference of the average absolute Gaussian curvature of the developing surface with different λ_L .

with different λ_L arrange from 0.05 to 0.2, which present slow to fast growth rate respectively. As shown in Fig.8, slower skull growth rates can generate more convolutions. A slower growth rate means smaller space for cortex to grow at the same time and thus more constraint force. By changing growth rate of dynamic skull, we can simulate condition with different constraint force - the slower the growth rate is, the higher the constraint force will be. This result further indicated that skull constraint is an important regulator of the cortical folding process.

4 Discussion and Conclusion

Our simulations demonstrate that mechanical constraints imposed by the skull are important regulators of cortical folding. However, it should be noted that our simulations indicate that skull constraint is not necessarily the dominant or initializing mechanism. Previous experimental observation has also shown that it is difficult to conclude that skull restraint initiates the development of cortical folding [11]. Our current model of neuronal growth also assumes isometric deformation. However, it should be noted that this isometric model is insufficient. In [12], it was demonstrated that there are systematic variations in absolute thickness and cell number across the cortical landscape, e.g., the gyral columns contain significantly more neurons than sulcal columns. Currently, the growth parameter for our simulations are either homogenous or set differently in manually selected regions. In future work we plan to infer or estimate growth parameters from MRI data of the developing brain by measuring cortical thickness or gray matter density. The cortical measurements will then be mapped to the simulation space via cortical surface registration algorithms allowing us to investigate how multiple, locally intrinsic and real differentiations influence the cortical folding process.

In this paper, we present a dynamic skull growth model for improved simulation of cerebral cortex folding. Our results show significant improvement by the dynamic model. In our current model, only neuronal growth and dynamic skull constraints are modeled and simulated. In future, we intend to examine, model and simulate more premises regarding cortex folding mechanisms. For instance, as previously stated in the tension-based theory of morphogenesis [13], cortical patterns are considered to be the result of minimization of global tension energy along axons, dendrites, and neuroglia. Our future work will extend the computational system to a more general framework that could easily plug-in new mechanical force models, implemented according to different biological premises such as the tension-based morphogenesis theory [13]. By using above approach, different computational models of cortical folding mechanisms could be independently developed, and could be easily integrated to study their interactions. After reaching such a stage, it would be possible to simulate more realistic cortical folding patterns that are similar to those of a real, human cerebral cortex.

References

1. Talairach, J., Tournoux, P.: Co-planar Stereotaxic Atlas of the Human Brain. Thieme, New York (1988)
2. Fischl, B., et al.: Cortical Folding Patterns and Predicting Cytoarchitecture. Cereb Cortex 18(8), 1973–1980 (2008)

3. Raghavan, R., et al.: A continuum mechanics-based model for cortical Growth. J. Theor. Biol. 187, 285–296 (1997)
4. Toro, R., Burnod, Y.: A Morphogenetic Model for the Development of Cortical Convolutions. Cerebral Cortex 15, 1900–1913 (2005)
5. Nie, J., et al.: A Computational Model of Cerebral Cortex Folding. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 458–465. Springer, Heidelberg (2009)
6. Le Gros Clark, W.: Deformation patterns on the cerebral cortex. In: Essays on Growth and Form, pp. 1–23. Oxford University Press, Oxford (1945)
7. Liu, T., et al.: Brain Tissue Segmentation Based on DTI Data. NeuroImage 38(1), 114–123 (2007)
8. Yeo, B.T.T., et al.: Shape Analysis with Overcomplete Spherical Wavelets. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G., et al. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 468–476. Springer, Heidelberg (2008)
9. Zhang, T., et al.: Parametric representation of cortical surface folding based on polynomials. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C., et al. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 184–191. Springer, Heidelberg (2009)
10. Liu, T., et al.: Deformable Registration of Cortical Structures via Hybrid Volumetric and Surface Warping. NeuroImage 22(4), 1790–1801 (2004)
11. Barron, D.: An experimental analysis of some factors involved in the development of the fissure pattern of the cerebral cortex. J. Exp. Zool. 113, 553–581 (1950)
12. Hilgetag, C.C., et al.: Developmental mechanics of the primate cerebral cortex. Anat. Embryol. 210, 411–417 (2005)
13. Van Essen, D.C.: A tension-based theory of morphogenesis and compact wiring in the central nervous system. Nature (1997)

Coupled Personalisation of Electrophysiology Models for Simulation of Induced Ischemic Ventricular Tachycardia

J. Relan[1], P. Chinchapatnam[2], M. Sermesant[1,2,⋆], K. Rhode[2], H. Delingette[1], R. Razavi[2], and N. Ayache[1]

[1] Asclepios Research Project, INRIA, Sophia Antipolis, France
[2] Division of Imaging Sciences, St. Thomas Hospital, King's College London, UK
maxime.sermesant@sophia.inria.fr

Abstract. Despite recent efforts in cardiac electrophysiology modelling, there is still a strong need to make macroscopic models usable in planning and assistance of the clinical procedures. This requires model personalisation i.e. estimation of patient-specific model parameters and computations compatible with clinical constraints. Fast macroscopic models allow a quick estimation of the tissue conductivity, but are often unreliable in prediction of arrhythmias. On the other side, complex biophysical models are quite expensive for the tissue conductivity estimation, but are well suited for arrhythmia predictions. Here we present a coupled personalisation framework, which combines the benefits of the two models. A fast Eikonal (EK) model is used to estimate the conductivity parameters, which are then used to set the parameters of a biophysical model, the Mitchell-Schaeffer (MS) model. Additional parameters related to Action Potential Duration (APD) and APD restitution curves for the tissue are estimated for the MS model. This framework is applied to a clinical dataset provided with an hybrid X-Ray/MR imaging on an ischemic patient. This personalised MS Model is then used for *in silico* simulation of clinical Ventricular Tachycardia (VT) stimulation protocol to predict the induction of VT. This proof of concept opens up possibilities of using VT induction modelling directly in the intervention room, in order to plan the radio-frequency ablation lines.

1 Introduction

Cardiac arrhythmias are increasingly being treated by Radio-Frequency (RF) ablation procedures. These procedures still have unsatisfactory success rates of only 30−60% for VT, due to non availability of clinical consensus on optimum RF ablation patterns [1]. There is still a need for substantial guidance in locating the optimum ablation lines. This guidance can be provided by personalised *in silico* cardiac electrophysiology models. Personalisation means estimation of the patient-specific model parameters which best fit the clinical data. It is required

⋆ Corresponding author.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 420–428, 2010.

to reveal hidden properties of the tissue and to develop predictive models that can be used to improve therapy planning and guidance. There are a variety of cardiac EP models developed at various scales. These models can be broadly categorised into three main categories: Biophysical Models (BM), Phenomenological Models (PM) and Eikonal Models (EM). BM [2] model ionic currents and are the most accurate and complex but hardly suitable for parameter estimation from clinical data. PM [3] are based on PDEs and mimic only the shape of action potential and are on intermediate level. EM [4] describes only the time at which a depolarisation wave reaches a given point. They can be very fast in computation [5], but less reliable in arrhythmia predictions due to the complexity of both refractoriness and curvature. To introduce models directly in the intervention room, the requirements are a low computational complexity, fast estimation of parameters and reliable predictions. These attributes cannot be found in one single model, thus here we present a new approach, where we combine two models to obtain these attributes and show an application to a clinical dataset. We also show how such personalised model can then be used to simulate *in silico* a clinical VT Stimulation protocol and can be potentially used to plan optimum RF ablation lines. In this paper, we present a coupled personalisation framework, which is fast and combines the benefits of an Eikonal (EK) model with those of a simplified biophysical model, the Mitchell-Schaeffer (MS) model. The fast 3D EK model is used to estimate the tissue conductivity parameter over the ventricles from the non-contact mapping of endocardial surface potential, using an adaptive iterative algorithm. This is then used to set the conductivity parameter of the 3D MS model. Additional parameters related to APD and APD restitution property of the tissue are then estimated locally using directly the 3D MS model. This framework is applied to a clinical data of an ischemic patient, containing of MR data for geometry and scar detection and electrophysiological data obtained from non-contact mapping. This data is obtained using Hybrid X- ray/magnetic resonance (XMR) suites [6]. The personalised 3D MS model is then used to simulate a clinical VT-Stim procedure to show a potential application of VT induction modelling.

2 Simulation of Electrophysiology Models

Anisotropic Fast Marching Model (EK Model): The EK model simulates the propagation of the depolarization wave in quiescent tissue, ignoring repolarisation phase. The EK model is governed by eikonal-diffusion (ED) equation and is based on Fast Marching Method (FMM), it can be written as

$$c_0\sqrt{D(x)}\left(\sqrt{\nabla T(x)^t\mathbf{M}\nabla T(x)}\right) - \nabla\cdot(D(x)\mathbf{M}\nabla T(x)) = \tau(x) \qquad (1)$$

where the superscript t denotes transpose, c_0 is a dimensionless constant, and $\tau(x)$ is the cell membrane time constant. $D(x)$ is the square of the tissue space constant along the fiber and is related to the specific conductivity of the tissue in the fiber direction. The anisotropicity is incorporated in the Diffusion tensor and

is given by **M**. The nonlinear term (Eq 1) is solved using a fixed point iterative method combined with a very fast eikonal solver as explained in [7].

Biophysical Model (MS Model): The MS model [8] is a 2-variable simplified biophysical model derived from the 3-variable Fenton Karma (FK) ionic model [9]. It models the transmembrane potential as the sum of a passive diffusive current and several active reactive currents including a sodium ion (influx) current and a potassium ion (outflux) current. Unlike FK model, it does not model the Calcium ion current. The MS model is described by the following system of Ordinary Differential Equations (ODE),

$$\begin{cases} \partial_t u = div(D\nabla u) + \frac{zu^2(1-u)}{\tau_{in}} - \frac{u}{\tau_{out}} + J_{stim}(t) \\ \partial_t z = \begin{cases} \frac{(1-z)}{\tau_{open}} \; if \; z < z_{gate} \\ \frac{-z}{\tau_{close}} \quad if \; z > z_{gate} \end{cases} \end{cases} \tag{2}$$

where, u is a normalised transmembrane potential variable, and z is a gating variable for sodium ion influx which makes the gate open and close, thus depicting the depolarisation and repolarisation phase. $J_{in} = (zu^2(1-u))/\tau_{in}$ represents the inward sodium ion current which raises the action potential voltage and $J_{out} = -u/\tau_{out}$ represents the outward potassium ion current that decreases the action potential voltage describing repolarisation. J_{stim} is the stimulation current, at the pacing location. The diffusion term in the model is controlled by the diffusion tensor D. This spatial diffusion can be related to a pseudo-conductivity. In the longitudinal direction of the fibre, this pseudo-conductivity is set to d which is one of the parameters we adjust, and to $d/2.5^2$ in the transverse directions. The electrophysiology model is solved spatially using P1 Finite Element Method (FEM), and in time using an semi-implicit scheme as Modified Crank-Nicolson/Adams-Bashforth (MCNAB) scheme, which is evaluated in terms of accuracy, stability and computational time [10].

3 Coupled Personalisation Method

Apparent Conductivity Parameter Estimation: Cardiac tissue conductivity is a crucial feature for the detection of conduction pathologies. The Apparent Conductivity (AC) of the tissue can be measured by the parameter $D(x)$ in the EK model. It is initially estimated on the endocardial surface as a global value using a simple bisection method which matches the average conduction velocity of the measured Depolarisation Time (DT) isochrones to the simulated ones. Using it as an initial guess, an adaptive multi-level zonal decomposition algorithm is used, which minimizes the mean-squared difference of the simulated and measured DT isochrones at each level using a Brent's Optimisation Algorithm presented in [5]. Due to the absence of transmural electrical propagation information, we assume no variation across the left ventricle myocardium (excluding LV endocardium and scars) and hence we prescribe a single value for the myocardial tissue across the LV wall. The AC values for RV endocardium

and RV myocardial mass are set at 5.0 mm and 0.64 mm (from literature [4]). The LV myocardial AC value is estimated by one-dimensional minimisation of the following cost function (mean-squared difference of simulated and measured isochrones at endocardium + squared difference of simulated and measured QRS duration). The simulated QRS duration is calculated as the difference between the maximum and the minimum depolarisation times in the biventricular mesh and the measured QRS duration is estimated from the surface ECG.

Coupling of EK and MS Model Parameters: The AC parameter for EK model d_{EK} ($D(x)$ in Eq 1) is a scale for the diffusion speed of the depolarisation wavefront in the tissue. The diffusion tensor used in Eq 1 is $\mathbf{M} = \mathbf{A}\bar{\mathbf{D}}\mathbf{A}^t$, where $\mathbf{A}$ is the matrix defining the fiber directions in the global coordinate system and $\bar{\mathbf{D}} = d_{EK}\text{diag}(1, \lambda^2, \lambda^2)$. λ is the anisotropic ratio of space constants transverse and along the fiber direction and is 0.4 in human myocardium [4]. The model Conduction Velocity (CV) is related to d_{EK} as,

$$c_{EK} = \frac{c_0\sqrt{d_{EK}}}{\tau} \text{ in 1D and } c_{EK} = \alpha_{EK}\sqrt{d_{EK}} + \beta_{EK} \text{ in 3D} \tag{3}$$

where the constants α_{EK} and β_{EK} are introduced to take into account the curvature effect and numerical diffusion and discretization errors in 3D. The corresponding conductivity parameter for MS model, d_{MS} is also a scale for the wave diffusion speed in the tissue. The diffusion tensor D used in Eq 2 is $D = \mathbf{A}\bar{\mathbb{D}}\mathbf{A}^t$, where $\mathbf{A}$ is the same as in EK model, but $\bar{\mathbb{D}} = d_{MS}\text{diag}(1, r, r)$ with r as conductivity anisotropy ratio in the transverse plane and is set to λ^2 as in EK model. The model CV here is related to d_{MS} as,

$$c_{MS} \propto \sqrt{\frac{d}{\tau_{in}}} \text{ in 1D and } c_{MS} = \alpha_{MS}\sqrt{d_{MS}} + \beta_{MS} \text{ in 3D} \tag{4}$$

where the constants α_{MS} and β_{MS} are introduced for the same reasons as of EK model, while τ_{in} is kept as a constant. The estimated AC parameter d_{EK} can then be used to estimate the parameter d_{MS}. The parameter d_{EK} gives model CV c_{EK}, which is similar to the actual measured data CV (c_{msd}) after the parameter estimation step. Thus to have MS model CV (c_{MS}) similar to the measured data, it has to be similar to EK model CV (c_{EK}). The constants α_{EK} and β_{EK} represent numerical errors for EK model based on FMM. They are different from the constants α_{MS} and β_{MS}, which is based on FEM. These constants are determined in 3D for a mesh representing the slab of a tissue ([0 10] $\times$ [0 10] $\times$ [0 10]) (with a mean edge length of tetrahedra same as the ventricular mesh). We performed several simulations with various d_{EK} and d_{MS} values and noted the corresponding c_{EK} and c_{MS} values. Then, we fit the analytical curves given in Eq 3 & 4 in least square sense and determine the constants. The constants estimated are $\alpha_{EK} = 5.21$, $\beta_{EK} = 0.07$, $\alpha_{MS} = 3.12$, $\beta_{MS} = 0.31$. Then, the personalised d_{MS} values are computed from corresponding estimated d_{EK} values using the condition, that $c_{msd} = c_{EK} = c_{MS}$ after personalisation.

Parameter Estimation for APD: APD for a single heart cycle is defined by the model as, $APD_{max} = \tau_{close} ln\,(1/h_{min})$ where $h_{min} = 4\,(\tau_{in}/\tau_{out})$ As

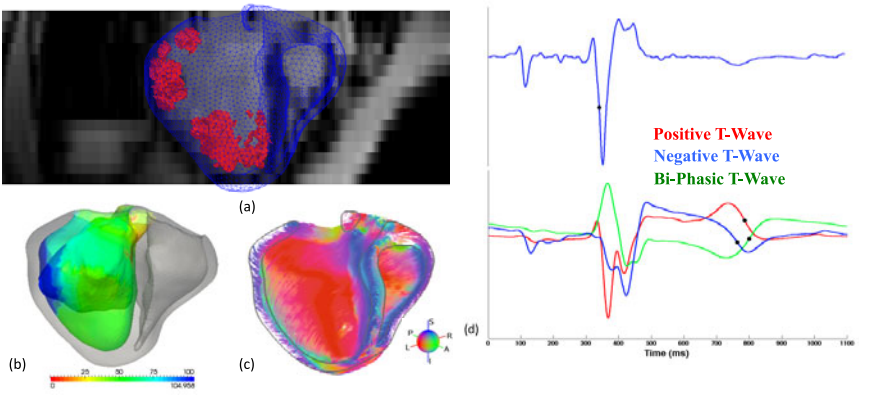


Fig. 1. (a) shows the MR data, segmented mesh with scars (in red), (b) shows XMR registration of Ensite LV surface with MR data mesh, with values projected from Ensite to MR LV surface,(c) shows the fibre orientation used, (d) shows the unipolar electrograms for detection (black dots) of depolarisation time (upper) and repolarisation time (lower) from positive(red), negative(blue) and biphasic(green) T waves

we have only one measured APD dependent on three parameters, We chose to estimate τ_{close}, while keeping the other parameter values from the literature [8]. The reason is that τ_{close} has no sensitivity towards the conductivity parameter estimation [10], whereas τ_{in} and τ_{out} do have. This defined relationship remains valid also in 3D thus allowing us to directly estimate locally at each vertex, the parameter τ_{close} without model simulations.To have a smooth gradation of APD from epicardium to endocardium, we diffuse the τ_{close} values spatially in the LV myocardium from Endocardium to Epicardium. For RV, we fix one value measured from the QT interval given through surface ECG.

Parameter Estimation for APD Restitution : APD Restitution is a property of a cardiac cell and defines its refractoriness. It is also heterogeneously distributed. It is a relationship between of the next cycle APD and the Diastolic Interval (DI) of the previous cycle. The slope of these restitution curves is controlled by τ_{open} and depicts the APD heterogeneity present at multiple heart rates. APD restitution curve for MS model is explicitly derived as, $f(DI_n) = \tau_{close} \ln\left(1 - (1 - h_{min})e^{-DI_n/\tau_{open}}/h_{min}\right)$, where $f(DI_n)$ is the succeeding APD and DI_n is the preceding DI at cycle n. Here, we use the τ_{close} values estimated as defined before, as it controls the APD at normal sinus rythm. And, the slope parameter, τ_{open} is estimated here with h_{min} fixed with values from the literature [8]. Here we minimise a cost function which minimises the error between model predicted APD ($f(DI_n)$) and actual measured APD (APD^{msd}_{n+1}) for a number of pacing frequency, where n is the cycle number. The Diastolic Interval (DI^{msd}) is measured from the data as $DI^{msd}_n = 1/f - APD^{msd}_n$, where f is the heart rate, detected from the ECG waveforms. The parameter optimisation method used here is a non-linear constrained Active-Set Algorithm, with constraints on τ_{open} to be in the range of literature values [8].

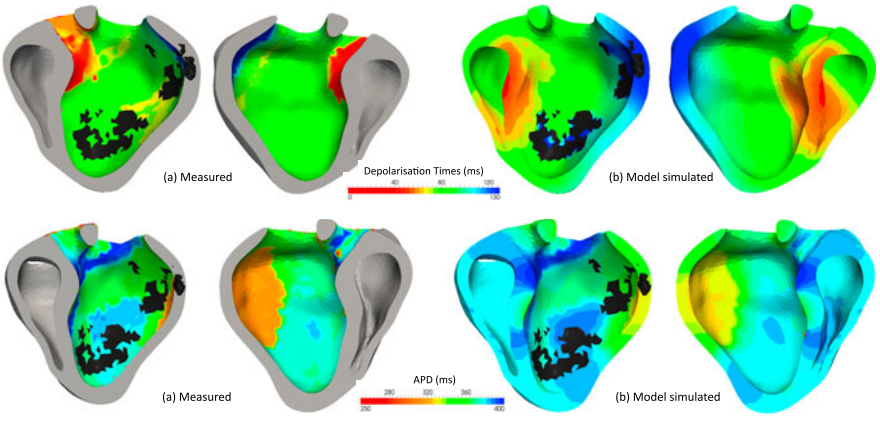


Fig. 2. Upper row shows the comparison of the measured Depolarisation Time (DT) isochrones on the LV surface only with model simulated DT isochrones on the whole heart, lower row shows the same for measured (LV surface only) and model simulated (whole heart) APD maps

4 Application to Clinical Data

The coupled personalisation framework is applied on a clinical data obtained during an electrophysiology study in the hybrid X-ray/MR environment. The electrical measurements obtained using the Ensite system were registered to the patient anatomy using XMR registration. The electrical data was collected with high pass filter settings for prominent QRS detection and with low pass filter for T Wave detection. The depolarisation times were detected from the dV/dt_{max} and d^2V/dt^2 of the unipolar electrograms V. And the repolarisation times were detected using dV/dt_{max} for the negative T wave, at the dV/dt_{min} for the positive T wave, and the mean time between dV/dt_{max} and dV/dt_{min} for the biphasic T waves. The data was collected from an ischemic patient at normal sinus rythm and 5 paced modes all at 100 beats per minute. The scars were segmented manually from the Delayed Enhancement MR data.

Estimated Parameters: The AC parameters estimated using EK model show a high conduction on the epicardium depicting the purkinje network and shows a conduction block near the scar as shown in Fig 3(a). The coupled MS model conductivity parameters are then estimated from AC. The mean absolute error on simulated depolarisation times with measured, after personalisation is $7.1ms$ for EK model and $8.5ms$ for MS model ($\approx 6-7\%$ of depolarisation duration ($131ms$)). The mean absolute error on APD is $8.71ms$ ($\approx 2\%$ of APD ($300ms$)), showing a good fit as well. Fig 3 (b) (white contour) shows the heterogeneity of the measured APD in terms of the estimated parameter τ_{close}, as shorter on the lateral wall of the LV compared to the septum. Also near the scar and the region between the two scars (called isthmus) Fig 3 (b) (black contour), we have

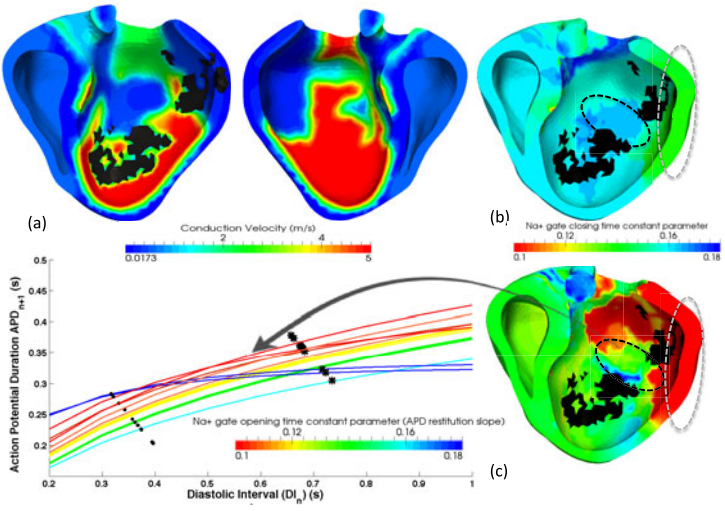


Fig. 3. (a) shows the conduction velocity estimated from AC maps, (b) shows the parameter τ_{close} estimated for APD, lower τ_{close} values has low measured APD (white contour) and vice versa, (c) shows the parameter τ_{open} estimated for APD restitution and the heterogeneity of the restitution curves for the isthmus (black contour), low τ_{open} values (red) have steep slopes & high (blue) have flat slopes for restitution curves

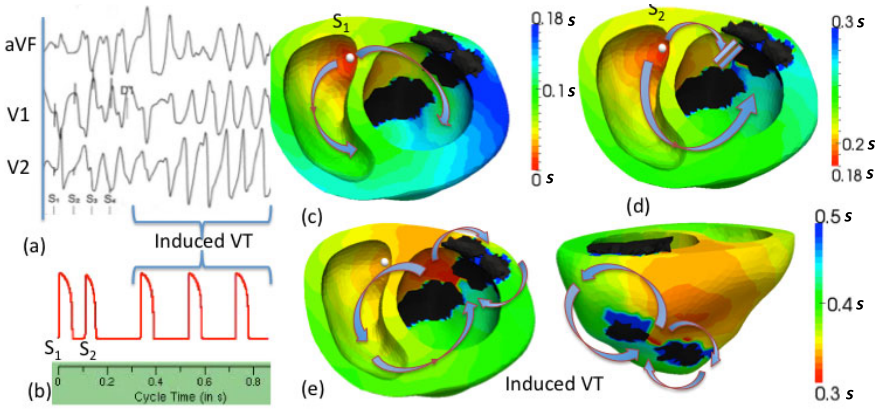


Fig. 4. (a) shows unipolar electrograms recorded for a clinical VT-Stim protocol, (b) shows the simulated protocol for two extrastimuli, with coupling interval of 100 ms. (c) show DT isochrones(in s) for S_1 stimulus and (d) shows for S_2, we have a unidirectional block created in the isthmus. (e) shows DT isochrones for induced monomorphic VT.

a longer APD compared to the neighbours. For the APD Restitution, the mean absolute error after fitting the resitution curves is $1.13ms$, showing a good fit also seen from the Fig 3(a). There is a APD restitution heterogeneity present for the lateral and septal walls as shown in Fig 3(c). Isthmus has flatter restitution

slopes compared to the neighbours, thus having a longer refractory period and causing a unidirectional re-entry as seen in VT-Stim procedure.

Ischemic Ventricular Tachycardia Stimulation: Programmed ventricular stimulation is a clinical protocol and consists of a number of extra stimuli introduced at two ventricular sites (RV-Apex & RV-Outflow tract), using various Cycle Lengths (CL), with varying coupling interval. This protocol is tested directly on the patient, without any planning, to collect information about the VT and to plan the RF ablation lines. It may be time consuming or fail, when VT is not inducible and recurrent. We use the personalised 3D MS model of the ischemic patient data to simulate *in silico* this protocol. Here we follow a conventional VT-Stim protocol with RV-Apex pacing site, 2 extrastimuli and a shortest coupling interval of 100 ms at 600 ms pacing cycle length. The results on inducibility are shown in Fig 4 and the causes of rentry were pacing location, restitution heterogeneity in isthmus compared to healthy Fig 3(c) and slow conductivity near the scars.

5 Conclusion

This novel approach of coupling models for fast estimation of hidden parameters related to the cardiac tissue such as conductivity, APD and APD restitution could enable the clinical use of cardiac electrophysiology models. The parameter estimation algorithm is used on a real interventional data and the obtained results are very encouraging. The estimated conductivity, APD and APD restitution parameters are able to identify the healthy areas from the pathological ones (scar and isthmus). The personalised MS model was able to simulate a clinical VT-Stim protocol in order to assess the risk of VT and fibrillation. This opens up possibilities of introducing patient-specific models in clinics to provide aid in treatment and planning of RF ablation procedures. In future, we need to evaluate the prediction ability of the personalised models for arrhythmias.

Acknowledgements. The research leading to these results was partially supported by a Prize from Microsoft Research Cambridge and by the EUHeart project (FP7/2007-2013 under grant agreement n 224495).

References

1. Aliot, E., Stevenson, W., Almendral-Garrote, J.: Ehra/hrs expert consensus on catheter ablation of ventricular arrhythmias. Europace (2009)
2. Noble, D.: Modeling the heart. Physiology (2004)
3. FitzHugh, R.: Impulses and physiological states in theoretical models of nerve membrane. Biophysical Journal (1961)
4. Keener, J., Sneyd, J.: Mathematical Physiology. Springer, Heidelberg (1998)
5. Chinchapatnam, P., Rhode, K., Ginks, M., Rinaldi, C., Lambiase, P., Razavi, R., Arridge, S., Sermesant, M.: Model-based imaging of cardiac apparent conductivity and local conduction velocity for diagnosis and planning of therapy. IEEE Trans. on Medical Imaging 27 (2008)

6. Rhode, K., Sermesant, M., Brogan, D., Hegde, S., Hipwell, J., Lambiase, P., Rosenthal, E., Bucknall, C., Qureshi, S., Gill, J., Razavi, R., Hill, D.: A system for real-time XMR guided cardiovascular intervention. IEEE TMI 24 (2005)
7. Sermesant, M., Konukoglu, E., Delingette, H., Coudiere, Y., Chinchaptanam, P., Rhode, K., Razavi, R., Ayache, N.: An anisotropic multi-front fast marching method for real-time simulation of cardiac electrophysiology. In: Sachse, F.B., Seemann, G. (eds.) FIHM 2007. LNCS, vol. 4466, pp. 160–169. Springer, Heidelberg (2007)
8. Mitchell, C.C., Schaeffer, D.G.: A two current model for the dynamics of cardiac membrane. B. of Mathematical Biology (2003)
9. Fenton, F., Karma, A.: Vortex dynamics in three-dimensional continuous myocardium with fiber rotation: Filament instability and fibrillation. In: CHAOS (1998)
10. Relan, J., Sermesant, M., Delingette, H., Pop, M., Wright, G.A., Ayache, N.: Quantitative comparison of two cardiac electrophysiology models using personalisation to optical and mr data. In: ISBI (2009)

Real Time Ultrasound Needle Image Simulation Using Multi-dimensional Interpolation

Mengchen Zhu[1] and Septimiu E. Salcudean[2]

[1] Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, GA, USA
mczhu@gatech.edu
[2] Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC, Canada
tims@ece.ubc.ca

Abstract. In this paper, we propose an interpolation-based method for simulating needle images in B-mode ultrasound. We parametrize the needle image as a function of needle position and orientation. We collect needle images under various spatial configurations in a water-tank using a guidance robot. Then we use multi-dimensional tensor-product interpolation to simulate images of needles with arbitrary poses and positions using the collected images. Interpolated needle images are superimposed on top of phantom image backgrounds. The similarity between the simulated and the real images is measured using a correlation metric. A comparison with in-vivo images is also performed. The simulation procedure is demonstrated using transverse needle images and extended to sagittal needle images and brachytherapy seed images. The proposed method could be used in clinical procedure training simulators.

1 Introduction

The simulation of needle insertion under ultrasound guidance would be useful in the planning and training of a number of clinical procedures. A needle insertion simulator would include a haptic interface to enable interaction with a deformable tissue model and a real time ultrasound image simulator to provide visual feedback. The image simulator requires both tissue and needle images to be rendered. Previous work (e.g. [7,13]) addressed primarily the simulation of B-mode images of soft tissue; the needle was simply rendered as a bright line in the longitudinal plane, without details or validation.

Needle images are highly *view-dependent*, with images changing drastically with the relative location between the needle and the ultrasound transducer. Physicians exploit this dynamic effect to judge the position of the needle, especially in the transverse plane [2]. For example, in prostate brachytherapy [5], physicians insert the needle until they see its tip, then rotate it to insure that the tip is at the desired transverse position. For simulation purposes, it is important to take into account the view-dependency property of needles.

In this paper, we propose to use interpolation on collected needle sample images to simulate new needle images in real time. The collected images are

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 429–436, 2010.

of various needle poses, so the simulator is able to reproduce view-dependent features of the needle images.

The classical way of simulating ultrasound images by interpolation is by slicing a 3D volume [8] and applying techniques such as voxel nearest-neighbor [9]. This approach cannot deal with view-dependent features of needle images. Indeed, looking at the needle from another angle is not equivalent to rotating the original image. Approaches based on physical acoustic models [4] are computationally expensive, while the faster methods using artificial acoustic models [7,13] lack accuracy. Hybrid methods [11] can be view-dependent and real-time, but they only simulate point scatterers which are not suitable models of needles, which have much more complex forms of scattering, such as the "comet tail" artifact. The idea of using images of an object under various poses and positions to simulate new object poses has been explored in the field of *image-based rendering* (e.g. [6]), but the ray-based parametrization from such references cannot be adopted to our case because there is no relative motion between the "camera" and the "light source" in our case – both are the ultrasound transducer.

2 Simulation Methods

It is noted that the image of the needle in water resembles that in the phantom/tissue. The basic idea of the simulation is therefore to first simulate the needle images in water by interpolating sample images collected in water and then superimpose the simulated images onto the tissue background.

We make the assumption that the needle image depends only on the local spatial configuration (position and orientation) of the intersection of the needle with the imaging plane. In this manner, we can represent the planar needle image as a function $I_n : \mathbb{R}^6 \rightarrow \mathbb{R}^2$ of the six degrees of freedom of the needle.

In the following, we first demonstrate our approach of parametrization, sample collection, interpolation, and validation on the transverse plane of a transrectal ultrasound (TRUS) probe. Using a slightly different parametrization, the same method could be applied to simulate images on the sagittal (longitudinal) plane and seed images.

2.1 Parametrization

The parametrization for the needle configuration with respect to the TRUS probe transverse plane is illustrated in Fig. 1a. Assuming the ultrasound beam profile has a constant thickness with respect to r along the Z (elevation) direction and the needle is inserted toward the negative Z direction, we define the imaging plane as the plane through which the needle enters the beam profile. This plane can be determined experimentally by inserting the needle until the tip starts to show up. We attach a Cartesian frame $\underline{C}_p$ at the probe origin Q_p in the imaging plane and another Cartesian frame $\underline{C}_n$ at the intersection of the imaging plane and the needle axis $\underline{P}_n$. Z_n coincides with the needle axis and Y_n indicates the bevel tip direction. We use (r, θ) to denote the polar coordinates of $\underline{P}_n$

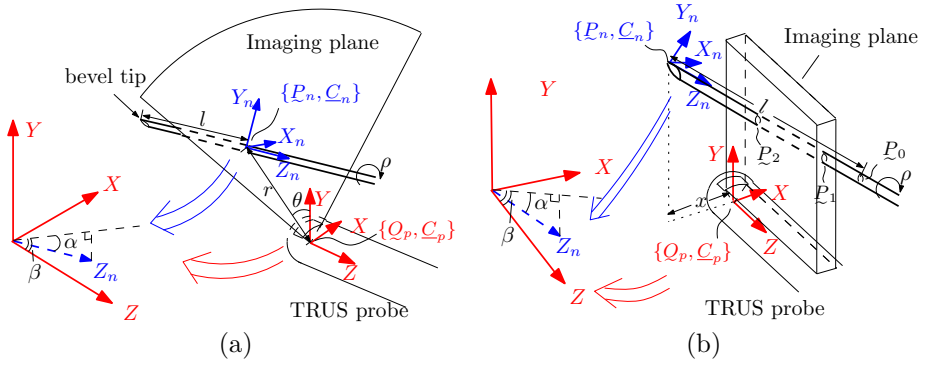


Fig. 1. Coordinate assignments for: (a) Transverse (b) Sagittal plane

in $\{\underset{\sim}{Q}_p, \underset{\sim}{C}_p\}$, and l to denote the needle depth. We describe the rotation transformation between $\underset{\sim}{C}_p$ and $\underset{\sim}{C}_n$ using roll-pitch-yaw angles (ρ, α, β). The needle image is assumed to be a function $I_n(r, \theta, l, \rho, \alpha, \beta)$ of these spatial variables. We have shown experimentally that the needle image is invariant with respect to θ. Indeed, if we keep other spatial parameters constant and vary θ, the image will only rotate by θ but will not change in intensity or shape otherwise:

$$I_n(r, \theta, l, \rho, \alpha, \beta) = R_\theta(I_n(r, 0, l, \rho_0, \alpha_0, \beta_0)) \tag{1}$$

where R_θ is a 2D rotation transform of the image with respect to the ultrasound probe center and (ρ, α, β) and $(\rho_0, \alpha_0, \beta_0)$ are related by the same rotation transform.This rotation invariance is due to the curvilinear geometry of the TRUS probe. To obtain images at an arbitrary θ, we rotate the image at $\theta = 0$. Thus the function to be sampled and interpolated is $I_n(r, 0, l, \rho_0, \alpha_0, \beta_0) : \mathbb{R}^5 \rightarrow \mathbb{R}^2$.

2.2 Tensor-Product Interpolation

Assuming that sample values of I_n are available through a collection method as described in the next section, the interpolation scheme seeks to estimate values of I_n at non-sample points. The speed of interpolation is of major concern. Nearest neighbor is the fastest interpolation method available, and is twice as fast as linear interpolation [12]. To strike a balance between speed and accuracy, we look for a scheme that is a hybrid of the nearest neighbor and linear interpolation. A suitable method is the *tensor-product interpolation* which enables the use of different piecewise polynomial interpolations – nearest neighbor and linear interpolations being special cases – along different dimensions [1]. The choice between the nearest neighbor and the linear interpolation for each dimension is based on the fact that during the needle insertion, the needle position on the imaging plane and its orientation are constrained by the tissue, therefore not likely to change much. It is the depth and roll angle that are under the physician's control and therefore responsible for the continuous changes of the

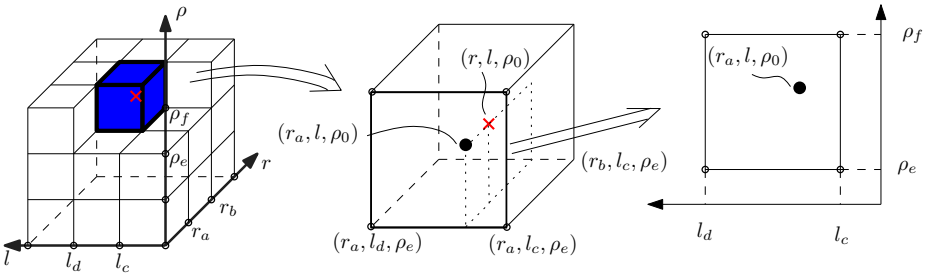


Fig. 2. Visualization of the interpolation process. Here only 3 dimensions are drawn. Each vertex represents a sample image. The red cross represents the queried configuration (r, l, ρ_0).

needle images. Using linear interpolation for l, ρ enables us to simulate these changes, and using nearest neighbor for other dimensions speeds up the interpolation considerably. Furthermore, using nearest neighbor for r will eliminate artifacts caused by averaging two images with features at different locations (e.g. the bright tail).

A general account of the tensor-product interpolation scheme is available in [1]. A key feature of the method is that changing the sequence of interpolation does not change the result, therefore we could perform nearest neighbor interpolation first to reduce the computational cost. Specifically we use nearest neighbor interpolation for r, α, β and linear interpolation for l, ρ. An operational procedure of the scheme applied to our problem is as follows and the key steps are visualized in Fig. 2:

1. Given the queried coordinate $(r, l, \rho, \alpha, \beta)$, calculate the rotated coordinate $(r, l, \rho_0, \alpha_0, \beta_0)$ in Eq. 1;
2. Find the enclosing hypercube in the 5-dimensional sample space that contains the rotated coordinate: $(r, l, \rho_0, \alpha_0, \beta_0) \in [r_a, r_b] \times [l_c, l_d] \times [\rho_e, \rho_f] \times [\alpha_g, \alpha_h] \times [\beta_i, \beta_j]$.
3. Find the nearest neighbors for r, α_0, β_0; say r_a, α_h, β_j respectively;
4. In the rectangular subspace $[l_c, l_d] \times [\rho_e, \rho_f]|_{(r_a, \alpha_h, \beta_j)} \in \mathbb{R}^2$, use linear interpolation to get the interpolant image $\tilde{I}_n$. This is just bi-linearly interpolating $I_n(r_a, l_c, \rho_e, \alpha_h, \beta_j)$, $I_n(r_a, l_c, \rho_f, \alpha_h, \beta_j)$, $I_n(r_a, l_d, \rho_e, \alpha_h, \beta_j)$, and $I_n(r_a, l_d, \rho_f, \alpha_h, \beta_j)$;
5. Now $\tilde{I}_n$ corresponds to a simulated needle image with the needle at coordinates $(r_a, 0)$ in the imaging plane. To simulate the image at (r, θ), we first translate $\tilde{I}_n$ by $(r_a - r)$ and then rotate by θ according to Eq. 1 to finally get the approximated $I_n(r, \theta, l, \rho, \alpha, \beta)$.

It can be shown that steps 1-4 of the above algorithm have a computational cost of $\mathcal{O}(N_p)$, where N_p is the number of pixels of the images being interpolated. In the current implementation $N_p = 80 \times 350$ and is the size of needle images cropped from the original 640×480 images (the rest is black background). The

speed of step 5 above largely depends on the interpolation method used for the spatial image rotations. In Matlab, using the nearest-neighbor will result in real time performance but may introduce aliasing artifacts or "jaggies". This was eliminated in an OpenGL implementation where linear interpolation can be computed in real time. On a PC with an Intel 2.2GHz dual core CPU, the overall algorithm can achieve an average frame rate of 44 FPS for a resolution of 640×480 in the Matlab implementation without major optimization efforts.

2.3 Sample Collection and Superimposition

All data collection was done in a water tank of size L×W×H = $50 \times 30 \times 20$ cm^3 filled with degassed water at 20℃. Brachytherapy needles (Bard®, 18 gauge) were imaged with the TRUS probe of an ultrasound machine (Ultrasonix SonixRP). The machine parameters were set to be the same as in a brachytherapy procedure, except for disabled Time Gain Control, single focus, and sound speed $c = c_{water}$. The time gain control was disabled during the data collection because in water there is very little attenuation with increasing depth; single focus was used because multiple focal points may sometimes cause split images in water. The samples were collected in the water tank rather than in a phantom because in water moving from one needle configuration to another leaves no needle tracks. Furthermore, the needle images in water are much more clearly defined.

A needle guidance robot previously developed [10] was used for the data collection. Since the robot does not move in l or ρ, we controlled these two variables by hand. The robot was programmed to remotely connect to the SonixRP and capture images automatically after each desired configuration was reached.

The range and sample numbers of each dimension are shown in the caption of Fig. 3. The ranges of α, β were kept small, as in a real brachytherapy procedure. Only positive rotation angles were collected due to symmetry.

The accuracy of the tensor-product interpolation depends on the spatial sampling rate. To ensure that the sampling rate is high enough, test images were first captured on a fine sample grid, then part of these images were used to interpolate the rest. Correlation was then used to compare the results, and the interpolated images were found to be similar to the real images when the sample numbers in Fig. 3 were used. As an example, for the roll angle ρ, when using images collected at every 180° to interpolate images of 45° ρ intervals, the correlation between the interpolated and the real images was 0.93; when using images collected at every 90°, the correlation was 0.97.

To simulate the needle images in the tissue, background images of a gelatine tissue mimicking phantom were collected, onto which the interpolated needle images in water were superimposed by addition. Superimposition by taking the maximum of the foreground and the background has also been tested, but the results looked less realistic. It was concluded that superimposition by addition gives satisfactory results. Note that the success of addition relies on the approximate linearity of the imaging process, which may be influenced by machine settings such as the time-gain-compensation.

3 Results and Validation

To validate the simulation method, 43 needle images in both the water tank and a homogeneous gelatine phantom with cellulose scatterers were collected using the same machine settings as used when the needle image samples were collected. The needle configurations were chosen to be different from those previously collected. Blank phantom background images were also collected before each insertion for use in the superimposition. Because images of the same modality are being compared, classical normalized cross-correlation is used as the similarity metric.

To begin with, we only measure the similarities between the areas adjacent to the needle in the images. This is justified because it is the needle image that is localized most of the time during the brachytherapy procedure and because the

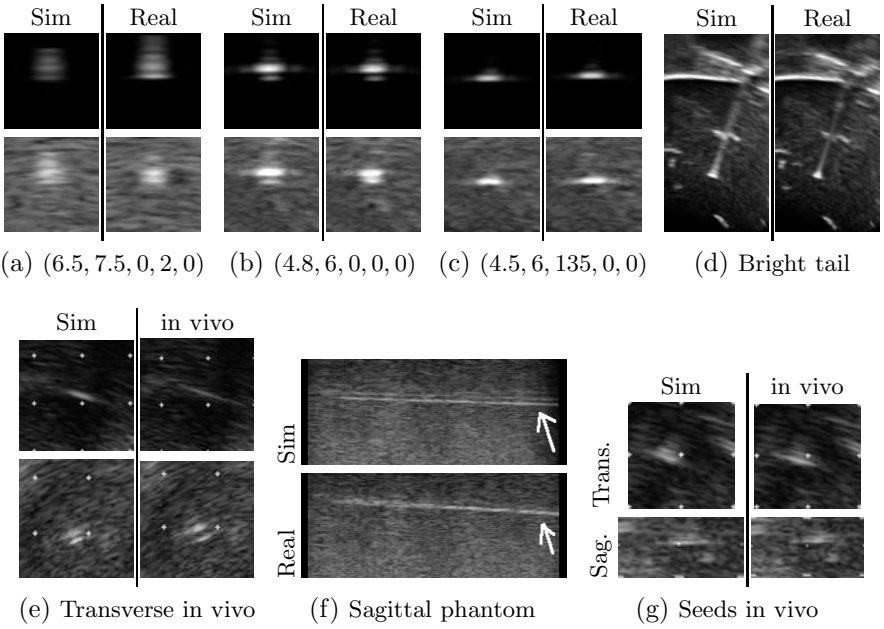


Fig. 3. (a) through **(c)**: comparison of simulated and real needle images of the transverse plane. **upper-left** quadrant: simulated image in water; **upper-right**: real image in water; **lower-left**: simulated image in phantom; **lower-right**: real image in phantom. Configurations are of the form $(r, l, \rho, \alpha, \beta)$. The range and sample number of variables are: r: 30 $\sim$ 65mm; 8 samples; l: 1 $\sim$ 22mm; 12 samples; ρ: 0 $\sim$ 180°; 3 samples; α: 0 $\sim$ 9°; 4 samples; β: 0 $\sim$ 15°; 3 samples; **(d)** Simulated and phantom needle image with bright tails; **(e)** Comparison of transverse simulated and in vivo needle images from a brachytherapy session; **(f)** Comparison of simulated and phantom sagittal needle images. The range and sample number of variables are: x: -0.7 $\sim$ 1.5cm; 23 samples; ρ: 0 $\sim$ 180°; 3 samples; α: 0 $\sim$ 6°; 3 samples β: 0 $\sim$ 15°; 4 samples; **(g)** Comparison of the simulated and in vivo seed images in the transverse (upper) and sagittal plane (lower).

background contributes to the false positive measure of the correlation score. For each of the 43 image pairs, an 80×80 region around the the needle image was cropped out and correlations were calculated. The mean correlation between the simulated and the real needle images in water is $cor_w = 0.86$ out of 1. The mean correlation between the simulated (after superimposition) and the real needle images in phantom is $cor_p = 0.79$.

In some occasions, a bright tail is visible distal to the needle image due to reverberation inside the needle. Our interpolation and superimposition method produces images similar to real images when the tail is present. The correlation between regions that contain the tail region and the phantom images has a typical score of 0.89, as shown in Fig. 3d.

The simulation results were also compared with several in vivo needle images collected during a brachytherapy session. To obtain a similar background for superimposition, the needles in the in vivo images were first erased by copying surrounding background using the image editing software GIMP (GNU Image Manipulation Program). Then the simulated needle images were superimposed. Example results are shown in Fig. 3e. Since there were no direct measurements of the needle position and orientation in-vivo, the simulated images were generated with the assumptions that the needle tip was in the imaging plane and that the needle was parallel to the Z axis (0 pitch and yaw angles). It is evident from Fig. 3e that under these conditions the simulated needle images are similar to the in vivo ones. The difference may be attributed to the different ultrasound machines being used for data collection – an Ultrasonix® machine was used for the sample image collection; a BK Medical® machine was used in the hospital.

Using slightly different parametrization and collection methods, the same method could be extended to simulate needle insertion in the sagittal plane and radioactive seeds in both planes in real-time (Fig. 1b sagittal plane coordinate assignment; Fig. 3f and 3g results)).

4 Discussion and Future Work

This paper presents a first view-dependent real-time needle image simulation method that is also suitable for simulating other wire-like objects such as radioactive seeds. The use of a robot to collect pose-dependent sample images proves to be accurate, efficient, and repeatable. The simulated images are not only visually appealing, but also accurate. The accuracy was validated objectively using a correlation metric and by comparison with in vivo data. The simulation method is compatible with existing real-time deformable tissue image simulations. Indeed, the needle simulation described here has recently been successfully incorporated into a real-time prostate brachytherapy simulator [3].It should be noted that the range of positions that our method can handle is only limited by the physical movement range of the brachytherapy robot used for data collection. Given more samples, the method can readily deal with a larger parameter range with little increased computational cost.

The simulation may potentially serve as a real-time “front-end” for a physically based simulator where sample images are simulated rather than collected.

Furthermore, the collected image sample database may potentially be used as training database for seed detection purposes.

In this paper, it was assumed that the needle does not change its pitch and yaw angles during the insertion. However, if the needle is not rigid, the pitch and yaw angles might be continuously changing when the needle is pushed sideways. Linear interpolation for these angles and piecewise rigid approximations may be needed to better simulate a bending needle.

References

1. Allen, M., Isaacson, E.: Numerical Analysis for Applied Science. Wiley Interscience, Hoboken (1998)
2. Chapman, G., Johnson, D., Bodenham, A.: Visualisation of needle position using ultrasonography. Anaesthesia 61(2), 148 (2006)
3. Goksel, O., Salcudean, S.: Haptic simulator for prostate brachytherapy with simulated ultrasound. In: Bello, F., Cotin, S. (eds.) Biomedical Simulation. LNCS, vol. 5958, pp. 150–159. Springer, Heidelberg (2010)
4. Jensen, J., Nikolov, I.: Fast simulation of ultrasound images. In: Ultrasonics Symposium, vol. 2. IEEE, Los Alamitos (2000)
5. Langley, S., Laing, R.: Prostate brachytherapy has come of age: a review of the technique and results. BJU International 89(3), 241–249 (2002)
6. Levoy, M., Hanrahan, P.: Light field rendering. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques, pp. 31–42. ACM, New York (1996)
7. Magee, D., Zhu, Y., Ratnalingam, R., Gardner, P., Kessel, D.: An augmented reality simulator for ultrasound guided needle placement training. Medical and Biological Engineering and Computing 45(10), 957–967 (2007)
8. Maul, H., Scharf, A., Baier, P., Wuestemann, M., Guenter, H., Gebauer, G., Sohn, C.: Ultrasound simulators: experience with the SonoTrainer and comparative review of other training systems. Ultrasound in Obstetrics and Gynecology 24(5), 581–585 (2004)
9. Rohling, R., Gee, A., Berman, L.: A comparison of freehand three-dimensional ultrasound reconstruction techniques. Med. Image Anal. 3(4), 339–359 (1999)
10. Salcudean, S., Prananta, T., Morris, W., Spadinger, I.: A robotic needle guide for prostate brachytherapy. In: IEEE International Conference on Robotics and Automation, ICRA 2008, pp. 2975–2981 (2008)
11. Shams, R., Hartley, R., 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)
12. Thevenaz, P., Blu, T., Unser, M.: Image interpolation and resampling. In: Handbook of Medical Imaging, Processing and Analysis, pp. 393–420 (2000)
13. Vidal, F., John, N., Healey, A., Gould, D.: Simulation of ultrasound guided needle puncture using patient specific data with 3D textures and volume haptics. Computer Animation and Virtual Worlds 19(2), 111 (2008)

Endoscopic Video Manifolds

Selen Atasoy[1,2], Diana Mateus[1], Joe Lallemand[1], Alexander Meining[3], Guang-Zhong Yang[2], and Nassir Navab[1]

[1] Chair for Computer Aided Medical Procedures (CAMP), TU Munich, Germany
{atasoy,mateus,lalleman,navab}@cs.tum.edu
[2] Visual Information Processing Group, Imperial College London, United Kingdom
{catasoy,gzy}@doc.ic.ac.uk
[3] Klinikum Rechts der Isar, TU Munich, Germany
alexander.meining@lrz.tu-muenchen.de

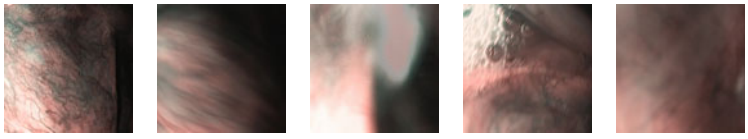
Abstract. Postprocedural analysis of gastrointestinal (GI) endoscopic videos is a difficult task because the videos often suffer from a large number of poor-quality frames due to the motion or out-of-focus blur, specular highlights and artefacts caused by turbid fluid inside the GI tract. Clinically, each frame of the video is examined individually by the endoscopic expert due to the lack of a suitable visualisation technique. In this work, we introduce a low dimensional representation of endoscopic videos based on a manifold learning approach. The introduced endoscopic video manifolds (EVMs) enable the clustering of poor-quality frames and grouping of different segments of the GI endoscopic video in an unsupervised manner to facilitate subsequent visual assessment. In this paper, we present two novel inter-frame similarity measures for manifold learning to create structured manifolds from complex endoscopic videos. Our experiments demonstrate that the proposed method yields high precision and recall values for uninformative frame detection (90.91% and 82.90%) and results in well-structured manifolds for scene clustering.

Keywords: Endoscopy, manifold learning, video segmentation, clustering.

1 Introduction

GI endoscopy is a widely used clinical technique for visualising the digestive tract. Current diagnosis and surveillance of GI diseases, ranging from Barrett's Oesophagus to oesophageal or colorectal cancer, are performed by visual assessment in GI endoscopy followed by necessary biopsies. Clinically, endoscopic videos also serve the postprocedural analysis performed by the expert and subsequent image processing for quantitative analysis. Currently, postprocedural analysis is typically performed by the endoscopic expert via visual assessment of each frame in the sequence. Such an analysis is complicated and time consuming mainly due to two reasons. First, in a typical endoscopic video sequence, there are usually a large number of poor-quality frames due to the blur caused by fast motion or out-of-focus imaging of the endoscope, specular highlights and artefacts caused by the turbid fluid inside the GI tract (Fig.1). Second, each frame

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 437–445, 2010.



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

Fig. 1. a) illustrates an ideal frame acquired by a state-of-the art GI endoscope. (b-e) show several challenges encountered in endoscopic videos; frames with motion blur b), specular highlights c), bubbles caused by the liquid inside the organ d) and blur caused by out-of-focus e).

in the sequence is inspected individually by the expert, as there exists no easily manageable visualisation technique for GI endoscopic videos.

In endoscopic video analysis, the focus is mainly directed towards detecting abnormalities [1,2,3] and uninformative frames [4,5]. These methods focus on defining specific features such as colour or texture and then detecting the frames containing them in order to present the expert only the detected informative frames instead of the whole content of the endoscopic video. Recently, representative frame extraction for content summary has also been investigated to aid the postprocedural analysis of wireless capsule endoscopy [6]. The aim of our work is to cluster the GI endoscopic videos in an *unsupervised* manner in order to allow the *expert* to easily eliminate or visualise only the parts of interest during postprocedural analysis. To this end, we introduce endoscopic video manifolds (EVMs); a low dimensional representation of endoscopic videos based on manifold learning that allows for clustering of different scenes as well as of poor quality frames.

Successful manifold learning algorithms have been proven to be beneficial for a range of image processing tasks, e.g. [7,8]. The main novelty of these methods in comparison to feature or intensity based image representation techniques lies in analysing a set of images based on their similarities. In [8], Pless proposed a video representation using a low dimensional image space and a trajectory for analysing natural video sequences. In this work, we will explore the use of manifold learning techniques to perform clustering on GI endoscopic videos.

The contribution of this work is twofold: firstly, from the medical point of view, we propose EVMs as a generic approach to cluster poor-quality frames as well as different segments of the GI endoscopic video in an unsupervised manner. This allows the experts to easily analyse the segment of interest. Secondly, in terms of theoretical contribution, we propose two inter-frame similarity measures for manifold learning, namely rotation invariant energy histograms and divergence of the optical flow field, which create structured manifolds from the complex endoscopic scenes. The first measure enhances the spectral differences between an ideal and a poor-quality frame while the second measure leads to closer localisation of similar frames on the manifold by considering temporal constraints among them. The design of these similarity measures is necessary as we are confronted with the difficult imaging conditions of endoscopy.

2 Methods

We address two tasks: clustering of *poor-quality frames* and *endoscopic scenes*. For each task our method creates a manifold representation using an appropriate inter-frame similarity measure and performs a clustering on the created EVM.

2.1 Overview of the Framework

An endoscopic video $\mathcal{I}$ can be represented by the set of its n individual frames $\{\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_n\}$. Each frame is a data point in the high dimensional input space $\mathcal{I}_1, \mathcal{I}_2, \cdots \mathcal{I}_n \in \mathbb{R}^{w \times h}$, where w and h are the width and height of the frames, respectively. Thus, the number of degrees of freedom (DoF) is equal to $w \times h$. However, due to the continuity of the video sequence, and therefore the large similarity between consecutive frames, the actual DoF is much smaller than this discrete representation enables. So, the high dimensional data points actually lie on a lower dimensional manifold $\mathcal{I}_1, \mathcal{I}_2, \cdots, \mathcal{I}_n \in \mathcal{M}$, where $\mathcal{M}$ is a manifold embedded in $\mathbb{R}^{w \times h}$. We compute the low dimensional EVM as follows:

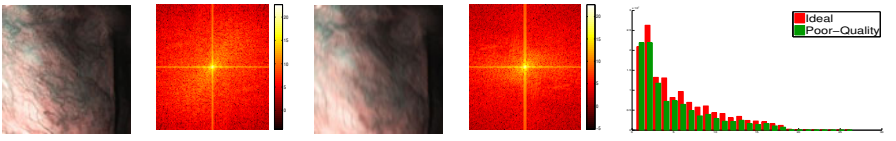
1. Defining the similarities: For each pair $(\mathcal{I}_i, \mathcal{I}_j)$, of the given n data points $i, j \in \{1, \cdots, n\}$, first a similarity measure is defined $W : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}$. W determines which images are considered to be similar and therefore kept as neighbours on the manifold. Thus, the similarity measure determines the structure of the manifold and should be designed carefully for each particular application. In the sections 2.2 and 2.3 we present the similarity measures designed for the addressed clustering tasks.

2. Computing the adjacency graph: Given the similarity matrix W, where the values $W(i, j)$ state the similarity between the frames $\mathcal{I}_i$ and $\mathcal{I}_j$, first, k-nearest neighbours of each data point are computed. Then, the adjacency graph is created as:

$$A(i,j) = \begin{cases} 1 & \text{if } i \in \mathcal{N}_j^k \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where $\mathcal{N}_j^k$ states the k-nearest neighbours of the j-th data point. Then, a connected component analysis is performed on the adjacency graph and the low dimensional manifold is computed for each component separately.

3. Learning the manifold: In this work, we use the *local* manifold learning based on Laplacian Eigenmaps (LE) [9]. The choice of this local method is driven by the observation that for the GI endoscopic videos distant data points on the manifold (corresponding to non-similar images) do not yield meaningful similarity measures. Therefore, local methods which do not take these similarities into consideration are better suited for our application compared to the global methods as used in [7,8]. To compute the LE, the eigenvalues and eigenvectors $\{f_1, \cdots, f_m\}$ of the Laplacian matrix $L = D - A$ are determined, where D represents the degree matrix $D(i,i) = \sum_j A(i,j)$. The m-dimensional ($m \ll w \times h$) representation of a frame $\mathcal{I}_i$ on the EVM is then given by $[f_1(i), \cdots, f_m(i)]^\top$.



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

Fig. 2. Rotation invariant energy histograms. a) and b) show an ideal frame and its power spectrum, respectively. c) and d) show a blurred frame and its power spectrum, respectively. e) shows energy histograms of the ideal and blurred frames.

4. Clustering on the manifold: Finally, the clustering of the uninformative frames and video segments is performed on the corresponding EVM using the K-means algorithm [10]. Thus, the endoscopic video $\mathcal{I}$ is represented as a set of l clusters $\mathcal{I} = \{C_1, \cdots, C_l\}$. The results of the clustering depends on the structure of the manifold and thus on the chosen similarity measure. As next we present, the construction of the EVMs for the two addressed tasks.

2.2 EVM for Clustering Uninformative Frames

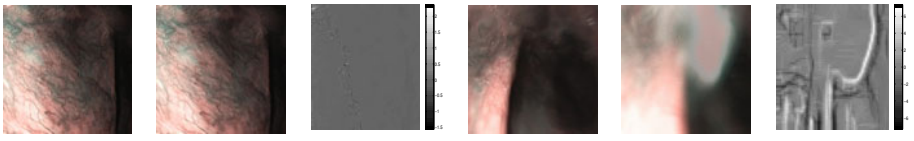
In order to create an EVM, where the poor-quality frames are closely localized, we propose to use a new inter-frame similarity measure based on the power spectrum of the images. In the frequency domain, the energy of an ideal frame is more distributed over low and high frequencies compared to a poor-quality frame whose energy is mainly accumulated only in low frequencies (Fig.2). Therefore, the EVM is created using the inter-frame similarity measure based on rotation invariant energy histograms. To this end, first the power spectrum of a frame $\mathcal{I}_i$ is represented in log-polar coordinates $\mathcal{F}_i(f, \theta)$, where f and θ state the frequencies and the orientations, respectively. Then the rotation invariant power spectrum is computed as: $\mathcal{F}_i(f) = \sum_\theta \mathcal{F}_i(f, \theta)$ and an histogram with B bins $\text{hist}(\mathcal{F}_i(f), B)$ is created. Finally, the EVM is created by using the following similarity measure for all pairs of frames $(\mathcal{I}_i, \mathcal{I}_j)$:

$$W_{\text{EH}}(\mathcal{I}_i, \mathcal{I}_j) = \pi - acos\left(\frac{\left\langle \text{hist}^b(\mathcal{F}_i(f), B), \text{hist}^b(\mathcal{F}_j(f), B)\right\rangle}{||\text{hist}^b(\mathcal{F}_i(f), B)|| \cdot ||\text{hist}^b(\mathcal{F}_j(f), B)||}\right), \qquad (2)$$

where hist^b states the b-th bin of the histogram, $\langle \cdot, \cdot \rangle$ is the dot product and $||hist||$ denotes the norm of the B-valued histogram vector. In this paper, we use $b = 30$ for our experiments. However, it is noted that there has not been a significant difference in the manifold structure for different values of n.

2.3 EVMs for Clustering Endoscopic Scenes

For clustering endoscopic scenes, we create two different EVMs; the first one based on the endoscope motion considering the temporal constraints (Sec.2.3a) and the second one considering the appearance similarities of all frames (Sec.2.3b).



(a) (b) (c) (d) (e) (f)

Fig. 3. Divergence of optical flow field. a) and b) show two consecutive frames with ideal conditions and c) illustrates the small and constant divergence field of the optical flow between a) and b). d) and e) show two consecutive frames with different conditions (one ideal and one non-informative frame). e) illustrates the varying divergence field with of the optical flow field between them.

a) Optical Flow Based EVMs: Changes in a GI-endoscopic video are caused mainly by the motion of the endosopce. Therefore a measure of the camera motion indicates directly a change in the observed scene. We propose using the optical flow divergence which measures the smoothness of camera motion field. This measure will lead to a high similarity between two images only if the scene and the imaging conditions (such as blur, specular highlights) are similar (Fig.3) If the optical flow field $\Phi_i^j(x, y)$ from i-th frame ($\mathcal{I}_i$) to j-th frame ($\mathcal{I}_j$) is a smooth motion field, then the divergence at each location will be close to 0. Thus, the similarity between $\mathcal{I}_i$ and $\mathcal{I}_j$ is computed as:

$$W_{\text{DOFF}}(\mathcal{I}_i, \mathcal{I}_j) = 1 - \frac{\psi_i^j}{\max(\psi_i^j)}, \qquad \psi_i^j = \sum_{x=1}^{w} \sum_{y=1}^{h} |\nabla \Phi_i^j(x, y)|, \qquad \Phi_i^j : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}^2 \tag{3}$$

where ∇ is the divergence operator $\nabla = \partial/\partial x + \partial/\partial y$. In order to consider temporal constraints, the k-nearest neighbours of a frame $\mathcal{I}_i$ are searched only within the frames $\{\mathcal{I}_{i-s}, \cdots, \mathcal{I}_{i+s}\}$, where s is the size of the search window (25 frames in this study). For the computation of the optical flow method, we use without loss of generality the optical flow method of Black and Anandan [11].

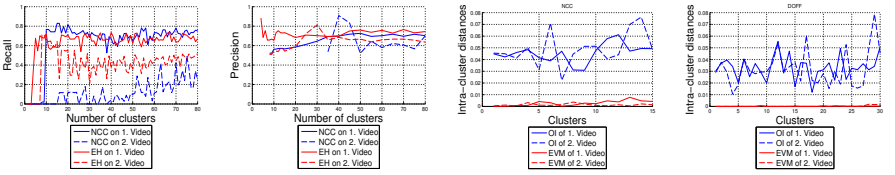
b) Intensity Based EVMs. Finally, we also create EVMs using the Normalised cross correlation (NCC) as similarity measure: $W_{\text{NCC}}(\mathcal{I}_i, \mathcal{I}_j) = NCC(\mathcal{I}_i, \mathcal{I}_j)$.

3 Experiments and Results

The experiments are conducted on two upper GI narrow-band endoscopic videos consisting of 1834 and 1695 frames. The datasets are acquired by an endoscopic expert at two different GI-endoscopic procedures. The ground truth labelling of poor-quality frames is performed manually by the expert for both videos.

3.1 Clustering Uninformative Frames

For this task, two EVMs are created using W_{EH} and W_{NCC} similarity measures. For quantitative analysis, *recall* and *precision* values of each clustering are



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

Fig. 4. a)Recall and b)precision values for clustering uninformative frames on EVMs created with W_{EH} and W_{NCC}. Note that for the 2. dataset W_{NCC} starts finding uninformative clusters only after 35 clusters, whereas W_{EH} shows a more stable performance. Intra-cluster distances of clustering on EVMs using c) W_{NCC} and d) W_{DOFF} as compared to clustering of original images using W_{NCC} similarity measure.

Table 1. Best recall and precision values of for clustering poor-quality frames

	Video 1			
	Max. Recall	Num. Clusters	Max. Precision	Num. Clusters
W_{NCC}	82.90%	15	74.13%	72
W_{EH}	73.71%	67	88.35%	4

	Video 2			
	Max. Recall	Num. Clusters	Max. Precision	Num. Clusters
W_{NCC}	74.13%	72	90.91 %	40
W_{EH}	71.21%	8	81.25%	30

evaluated over the number of clusters from 1 to 80. After clustering on the EVMs, clusters with more than 50% uninformative frames are labelled as uninformative. Particularly for this task, W_{EH} yields nicely structured manifolds, where informative and uninformative frames are well separated as shown in Fig.5(a1)-(a3). This is also reflected in the recall-precision curves (Fig.4a-b), where using 7 clusters on this EVM one can cluster apart 70.16% of all uninformative frames (recall) with a precision of 65.61%. Best recall and precision values are summarized in Table 1.

3.2 Clustering Endoscopic Scenes

The clustering of different segments is performed on optical flow (Sec.2.3a) and intensity based EVMs (Sec.2.3b). Inclusion of temporal constraints for optical flow based EVM requires the use of a larger number of clusters. Therefore, the optical flow and the intensity based EVMs are clustered using 30 and 15 clusters, respectively. The results are compared to K-means clustering perfomed on the original images using the same number of clusters (15 and 30) and W_{NCC} similarity measure. (Fig.5b,c) show the EVMs and examples of the clustered frames. For quantitive evaluation normalized intracluster distances (icd) are measured for all clusters C_i : $\mathrm{icd}(C_i) = \frac{\sum_{x \in C_i} \mathbf{x}-\bar{\mathbf{x}}_\mathbf{i}}{\max_{x \in C_i} \mathbf{x}-\bar{\mathbf{x}}_\mathbf{i}}$, where $\bar{\mathbf{x}}_\mathbf{i}$ denotes the centre of cluster C_i. Fig.4c,d show the decrease in icd when using the W_{DOFF} and W_{NCC} manifolds. This implies that the proposed similarity measures lead to structured manifolds that allow for better separability of the clusters. We further evaluate

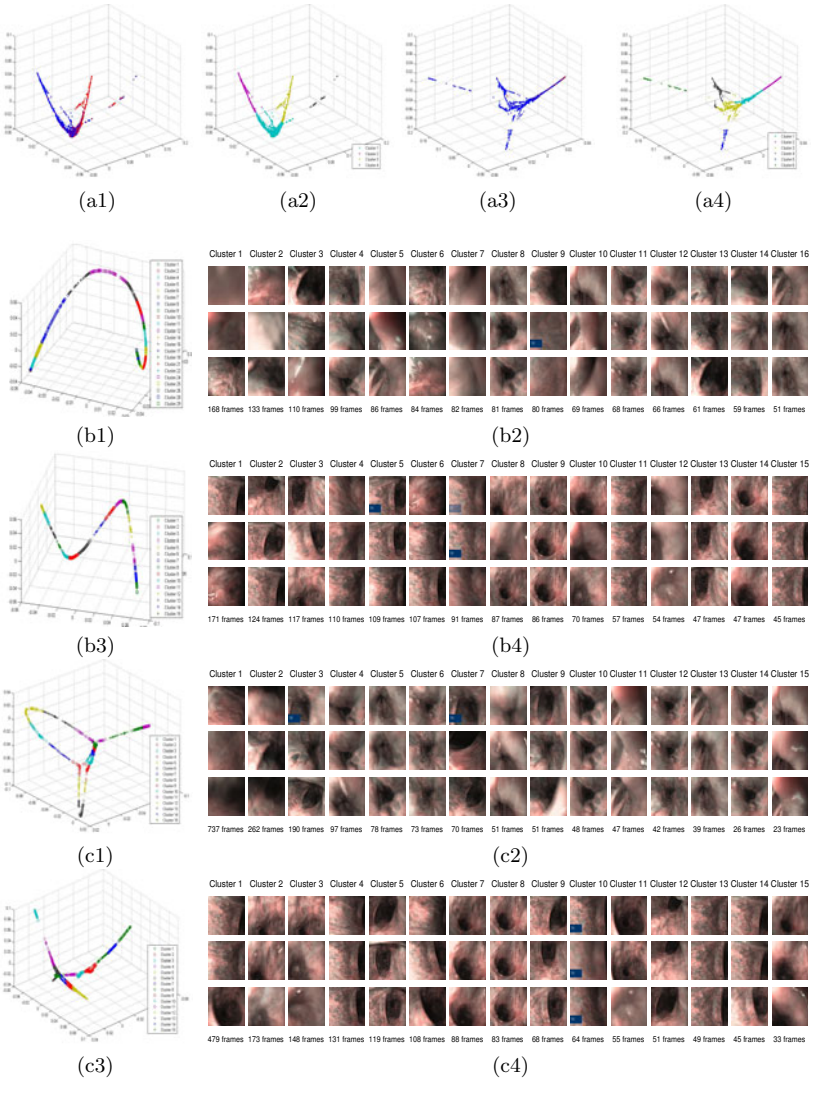


Fig. 5. (a1) and (a3) show the 3-dimensional EVMs of 1. and 2. endoscopic video, respectively. The red points illustrate the poor-quality frames in the ground truth labelling. (a2) and (a4) show the clustering results on the EVMs for the 1. and 2. video, respectively. The use of W_{EH} in manifold learning leads to structured EVMs where the poor-quality frames are clustered together. (b1) and (b3) Largest connected component of 3-dimensional EVM created using W_{DOFF} (Section 2.3) and the clustering on these EVMs for the 1. and 2. dataset, respectively. (b2) and (b4) show 15 example clusters for the 1. and 2. video; each column correspond to one cluster, where the rows show the first, center and the last frames of each cluster, respectively. (c1) and (c3) 3-dimensional EVM created using W_{NCC} (Section 2.3) and the clustering on these EVMs for the 1. and 2. dataset, respectively. (c2) and (c4) show clustering results using 15 clusters for the 1. and 2. video, respectively.

our results against manual labelling, where contiguous informative frames are labelled to be in the same cluster. The correlation between the ground truth and EVM clusterings is measured by the normalized mutual information, which is independent of the number of clusters. Clustering on W_{NCC} EVM yields 84.77% and 76.37%. Better results, 87.31% and 75.11%, are obtained with the proposed optical flow based clustering.

4 Conclusion

In this paper, we have proposed an effective framework for clustering endoscopic videos using EVMs. Key technical contribution of the paper includes: 1) we have addressed the task of clustering uninformative frames and endoscopic scenes from a different point of view than the methods in the literature, namely within a generic framework using the inter-frame similarities in an unsupervised manner. Our method provides a compact visualisation of the endoscopic video for subsequent analysis. 2) we have introduced two inter-frame similarity measures for manifold learning, namely rotation invariant energy histograms and divergence of optical flow field. Our experiments demonstrate that the proposed similarity measures yield well structured manifolds and thus lead to accurate clustering. The mathematical framework behind manifold learning has the particular advantage of being extendable by definition of the similarity measures. Therefore, even if particular characteristics of the imaging system changes, EVMs can be easily adopted by changing only the similarity measure.

References

1. Iakovidis, D., Maroulis, D., Karkanis, S.: An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy. Computers in Biology and Medicine 36, 1084–1103 (2006)
2. Hiremath, P., Dhandra, B., Hegadi, R., Rajput, G.: Abnormality detection in endoscopic images using color segmentation and curvature computation. Neural Information Processing, 834–841 (2004)
3. Li, P., Chan, K., Krishnan, S.: Learning a multi-size patch-based hybrid kernel machine ensemble for abnormal region detection in colonoscopic images. In: CVPR, vol. 2, p. 670 (2005)
4. Oh, J., Hwang, S., Lee, J., Tavanapong, W., Wong, J., de Groen, P.: Informative frame classification for endoscopy video. Medical Image Analysis 11 (2007)
5. Bashar, M., Kitasaka, T., Suenaga, Y., Mekada, Y., Mori, K.: Automatic Detection of Informative Frames from Wireless Capsule Endoscopy Images. Medical Image Analysis (2010)
6. Iakovidis, D., Tsevas, S., Polydorou, A.: Reduction of capsule endoscopy reading times by unsupervised image mining. Comp. Medical Imaging and Graphics (2009)
7. Balasubramanian, M., Schwartz, E., Tenenbaum, J., de Silva, V., Langford, J.: The isomap algorithm and topological stability. Science 295, 7 (2002)

8. Pless, R.: Image spaces and video trajectories: Using isomap to explore video sequences. In: ICCV, pp. 1433 (2003)
9. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural computation 15(6), 1373–1396 (2003)
10. Hartigan, J., Wong, M.: A k-means clustering algorithm. Appl. Stat. 28 (1979)
11. Black, M.J., Anandan, P.: A framework for the robust estimation of optical flow. In: ICCV, vol. 93, pp. 231–236 (1993)

Automated Training Data Generation for Microscopy Focus Classification

Dashan Gao[1], Dirk Padfield[1], Jens Rittscher[1], and Richard McKay[2]

[1] GE Global Research, One Research Circle, Niskayuna, NY, 12309
{gaoda,padfield,rittsche}@ge.com
[2] Omnyx, 800 Centennial Avenue, Building 4, 2nd Floor, Piscataway, NJ 08855
Richard.McKay@omnyx.com

Abstract. Image focus quality is of utmost importance in digital microscopes because the pathologist cannot accurately characterize the tissue state without focused images. We propose to train a classifier to measure the focus quality of microscopy scans based on an extensive set of image features. However, classifiers rely heavily on the quality and quantity of the training data, and collecting annotated data is tedious and expensive. We therefore propose a new method to automatically generate large amounts of training data using image stacks. Our experiments demonstrate that a classifier trained with the image stacks performs comparably with one trained with manually annotated data. The classifier is able to accurately detect out-of-focus regions, provide focus quality feedback to the user, and identify potential problems of the microscopy design.

1 Introduction

The adoption of digital microscopes in pathology studies enables the digitization of large tissue slides, which can then be stored, transferred, and analyzed electronically. As a crucial component of a digital microscope, the quality of the autofocus system directly determines the quality of the tissue scans. If the image is unfocused, the user will not be able to accurately characterize the tissue state as either diseased or healthy, and poor focus quality dampens the user's experience and can cause the user to loose confidence in the system. There is therefore a demand for developing focus quality assessment tools that detect such focus errors in whole slide images (WSI). Such ability to automatically measure image focus quality provides feedback to the pathologist as to the quality of the acquisition, and when such measures are fed back to the microscope itself, the system is able to correct problems even before the pathologist sees it.

Many approaches have been developed previously for predicting the focus plane for tissue imaging. Most of these approaches calculate some features on the image taken at a given microscope depth. For example, Liron *et al.* [1] measured the intensity of the reflected red laser light, Santos *et al.* [2] used various image gradient and histogram based metrics, Firestone *et al.* [3] compared measurements of the entropy and the range of the image intensity, and the authors

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 446–453, 2010.

of [4,5] adopted the normalized variance of image intensity. Other features include the sum of the absolute value of the image Laplacian [6] and the integrated amplitude modulus [7]. After the features are measured at one depth, the microscope depth is shifted, and the features are calculated again. If calculations are made at a sufficient number of depth locations and these locations span the actual location of the optimal focus, they will be sufficient to calculate the correct focus plane [8,9]. The requirement to acquire multiple images through the true focal plane can be prohibitively time consuming, especially for a high-throughput system. Furthermore, when the image quality measure is made on images after the acquisition step, the intermediate images may not be available, so the quality must be assessed on a single image.

To enable focus quality measurement on a single image, we trained classifiers to divide the image into in-focus and out-of-focus regions. This framework is considered a "no reference" approach since it enables the measurement of focus quality without the need for other images. Classifiers rely heavily on the quality and quantity of the training data, which is expensive and tedious for pathologists to generate, so we introduce a method of automatically generating training data using image stacks. Our quantitative experiments on annotated dataset and image stacks demonstrate that the classifiers can robustly separate in-focus from out-of-focus regions, provide useful feedback to the user about the focus quality, and even identify issues in the microscopy design such as stage tilt.

2 Region Classification for Focus Quality Assessment

We treat the image focus quality assessment as a binary classification problem, wherein the goal is to classify each WSI as either in-focus or out-of-focus. Guidance from pathologists indicate that a useful measure of focus quality is to divide the WSI into regions, classify each region as in-focus or out-of-focus, and calculate the percentage of in-focus regions. Some advantages of classifying smaller regions are that the image slide may be slightly tilted so that the focus depth varies slowly across the image, and it is likely that only part of the entire image is out-of-focus since the microscope does not fail completely for a scan in many cases. Therefore, we evenly divide each WSI into smaller regions and classify each region independently.

For classification, we represent an image region I by a vector of features, the choice of which is an important step to avoid features that bias away from the true focus [10]. We therefore utilize an extensive set of 44 features, most of which have been studied or suggested in previous microscope auto-focusing works [10,3,11]. The definitions and brief descriptions of these features are given in Table 1. We then adopt an AdaBoost classifier, which combines the outputs of a set of sequentially selected weak classifiers into a strong classifier to minimize the overall classification risk [12,13]. In our implementation, the weak classifier is generated by the linear discriminant analysis [14].

In general, the samples for training the classifier are manually cropped from images through careful visual inspection by expert pathologists. However, we

Table 1. Feature equations and descriptions. x and y are image coordinates. M and N are the width and the height of the image, respectively.

Neighborhood contrast features (24) Mean, variance, skew, and kurtosis of contrast calculated for each neighborhood Neighborhood $\mathcal{N}$ sizes: 5x5, 9x9, and 17x17	
Neighborhood ($\mathcal{N}$) contrast	$\max_{(x,y)\in\mathcal{N}} I(x,y) - \min_{(x,y)\in\mathcal{N}} I(x,y)$
Normalized neighborhood ($\mathcal{N}$) contrast	$\frac{\max_{(x,y)\in\mathcal{N}} I(x,y) - \min_{(x,y)\in\mathcal{N}} I(x,y)}{\max_{(x,y)\in\mathcal{N}} I(x,y)}$
Gradient and Laplacian features (3) S_x and S_y are the Sobel operators. L_x and L_y are Laplacian filter output in x and y directions.	
Tenenbaum Gradient [5,15]	$\frac{1}{MN}\sum_x\sum_y S_x(x,y)^2 + S_y(x,y)^2$
Laplacian sum [6]	$\frac{1}{MN}\sum_x\sum_y \lvert L_x(x,y)\rvert + \lvert L_y(x,y)\rvert$
Laplacian energy [16]	$\frac{1}{MN}\sum_x\sum_y (L_x(x,y) + L_y(x,y))^2$
Local image statistics features (7) $\mu = \frac{1}{MN}\sum_x\sum_y I(x,y)$ is the average of image intensity. p_i is the probability of observing a pixel of intensity i in the image.	
Intensity variance [5]	$\frac{1}{MN}\sum_x\sum_y (I(x,y)-\mu)^2$
Normalized intensity variance [5]	$\frac{1}{MN\mu}\sum_x\sum_y (I(x,y)-\mu)^2$
Auto correlation (x) [17]	$\frac{1}{MN}\sum_x\sum_y I(x,y)(I(x+1,y) - I(x+2,y))$
Auto correlation (y) [17]	$\frac{1}{MN}\sum_x\sum_y I(x,y)(I(x,y+1) - I(x,y+2))$
Standard deviation correlation (x) [17]	$\frac{1}{MN}\sum_x\sum_y I(x,y)I(x+1,y) - \mu^2$
Standard deviation correlation (y) [17]	$\frac{1}{MN}\sum_x\sum_y I(x,y)I(x,y+1) - \mu^2$
Shannon entropy [3]	$-\sum_i p_i \log_2(p_i)$
Wavelet features (10) Image decomposed into approximation A and detail D images. W_k is a wavelet channel with $k \in 0, 1, 2$ E_a is the energy of A, $E_d(k)$ is the energy of the k_{th} detail image α_k and β_k are the parameters for the generalized Gaussian distribution (GGD) representation for W_k	
Channel energy	$\frac{1}{MN}\sum_x\sum_y \lvert W_k(x,y)\rvert$
Defocus ratio [11]	$\frac{E_a}{E_d(1) + E_d(2)}$
Wavelet GGD parameters [18]	$P(W_k(x,y)) = \frac{\beta_k}{2\alpha_k\Gamma(1/\beta_k)} \exp\left\{-\left(\frac{\lvert W_k(x,y)\rvert}{\alpha_k}\right)^{\beta}_k\right\}$

will demonstrate that this tedious and slow process can be expedited by a clever utilization of tissue image stacks. Finally, given the in-focus/out-of-focus assignments from the classifier for all regions in a WSI, the focus quality is measured as the percentage of the regions that are out-of-focus. In addition, if there is time for re-acquisition, the images of the worst focus can be re-acquired.

3 Automatic Training Data from Image Stacks

To generate training data, images were manually collected from various tissue scans by two pathologists. The dataset consisted of image patches at two different magnification levels: 20× and 40×. For each magnification level, they labeled approximately 30 in-focus and 30 out-of-focus regions with the average size around 200 × 200 pixels. We extracted the 44 features from Table 1 for every image region covered by a 128 × 128 window, placed at intervals of 16 pixels in both x and y directions. Although this dataset was annotated by experts, a drawback is that its size is limited because the process of selecting such regions

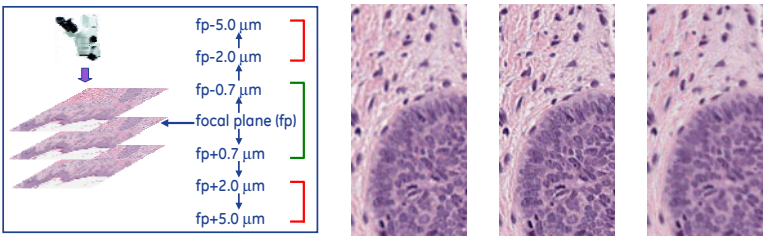


Fig. 1. Illustration of the generation of training data from image stacks. Images within $0.7\mu m$ of the focal plane are considered in-focus, and images between $2.0\mu m$ and $5.0\mu m$ away from the focal plane are considered out-of-focus.

is tedious and time-consuming. Trained with a small number of samples, the generalization of the classifier is limited.

To overcome the limited amount of data, we developed a method of automatically generating large amounts of training samples using image stacks. Image stacks are a set of images of the same tissue taken at different depths, with the focal plane manually determined. Figure 1 shows the general setup. According to pathologists, they consider a small range of slices around the focal plane to be in focus, and a range of slices far from the focal plane to be out-of-focus. Based on experience, they suggested the following thresholds: between $-0.7\mu m$ and $0.7\mu m$ was selected as "in-focus", and between $2.0\mu m$ and $5.0\mu m$ and between $-5.0\mu m$ and $-2.0\mu m$ were selected as "out-of-focus". The images between $0.7\mu m$ and $2.0\mu m$ and between $-2.0\mu m$ and $-0.7\mu m$ were determined to be too subtle to be accurately measured, and the images beyond $\pm 5.0\mu m$ are too out-of-focus to be useful in practice.

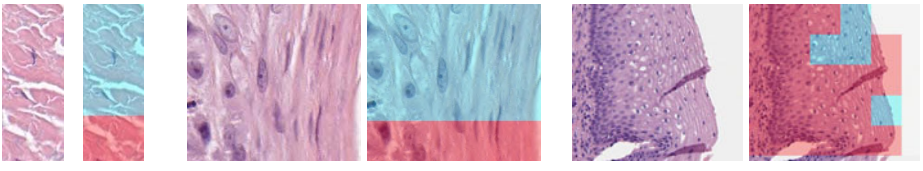
Using these thresholds, we generated an expanded image set that consisted of 37 tissue stacks, each of which includes 77 images scanned from $-5\mu m$ to $-2\mu m$, $-0.7\mu m$ to $0.7\mu m$, and $2\mu m$ to $5\mu m$ with an incremental step of $0.1\mu m$. The focus plane of each stack was carefully selected and verified by an expert before and after scanning. Each microscope image had a size of 2048×2048 pixels and was evenly divided into 256 non-overlapping samples each of size 128×128. Altogether, this led to 142,080 positive samples and 587,264 negative samples.

4 Experiments and Results

We have tested various parameters, such as the slide window size (128 or 256), and the initial weights for positive and negative samples, but we did not observe significant changes in the classification accuracy, which suggests the robustness of the classifier. For the results presented in the following, we chose a window size of 128 pixels, and equal initial weights for positive and negative samples. Table 2 provides a comparison of the classification accuracy of the AdaBoost classifiers trained with either the manually labeled samples or the image stacks.

Table 2. The average training and testing accuracy for a 10-fold cross validation on the manually labeled dataset (ManualSet) and the automatically generated dataset (StackSet). Both sets had high accuracy, but StackSet achieved better results than ManualSet, especially for 40× magnification. The training accuracy is the classification accuracy when testing on the training data (which gives an upper bound on the accuracy), and the testing accuracy represents the result on new test data. The test accuracy is close to the training accuracy, indicating no over-fitting during training.

	ManualSet-100iter.		ManualSet-30iter.		StackSet-30iter.	
	20X	40X	20X	40X	20X	40X
Training Accuracy (%)	91.939	87.062	91.918	86.319	92.806	93.421
Testing Accuracy (%)	90.854	86.679	90.996	85.683	92.772	93.371

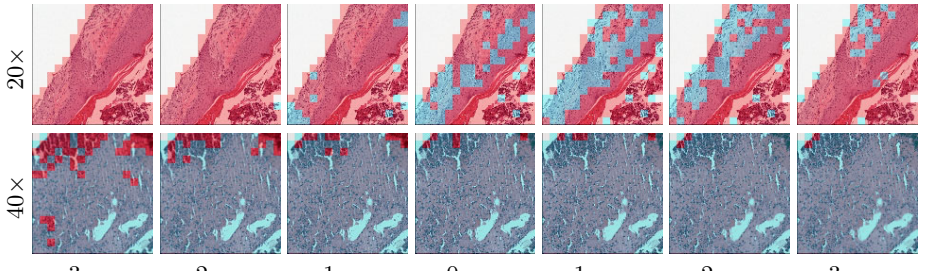


(a) (b) (c)

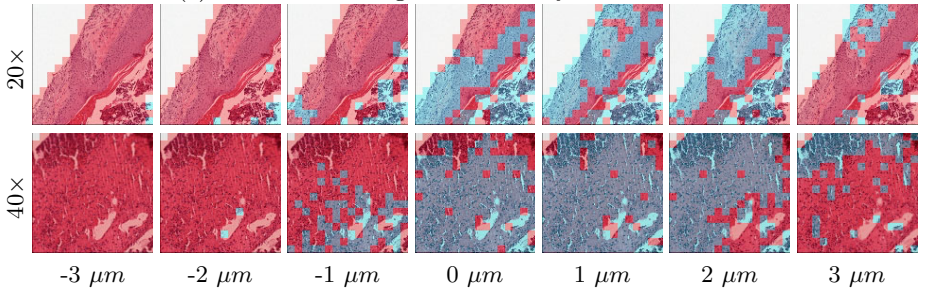
Fig. 2. Illustration of the sensitivity of the classifier trained with stacks. The classification results were overlaid on the images which were labeled by pathologists as (a,b) "in-focus" and (c) "out-of-focus". The overlaid blue and red colors represent in-focus and out-of-focus assignments, respectively. Closer inspection indicated that the focus quality indeed varies throughout the tissue, and the classifier were able to detect this.

The average training and testing accuracy are from a 10-fold cross validation, where we randomly, but evenly, divided the samples into ten subsets, and in each round, we chose nine of them for training and one for testing. For each classifier, 30 boosting iterations were considered, and we also show 100 boosting iterations for the manual set. In addition, two different magnification levels of 20× and 40× were trained separately because the characteristics of tissue images under these magnifications are significantly different.

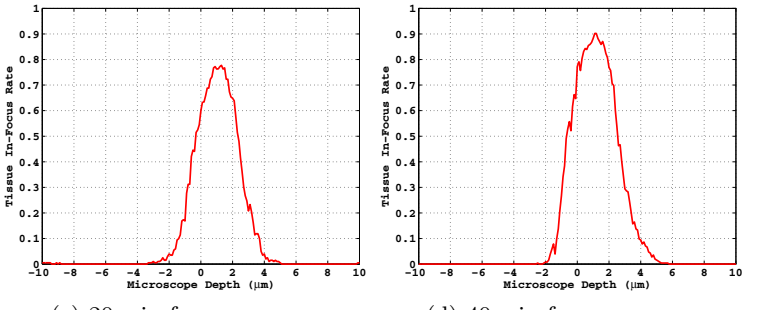
The table indicates the advantage of the proposed automatic data generation method. First, the classifiers obtained using the automatically generated data (StackSet-30iter) achieved better accuracy than those trained with the manual data for both 20× and 40× images. In particular, for 40× images the improvement is nearly 8%, indicating the usefulness of the vast amount of training samples provided by the image stacks method. Second, while the performance gap between 20× and 40× images for the manual data indicated that the 40× class is more difficult than the 20× class, the gap was bridged by using the stacks. Third, the classification accuracy does not improve much by increasing the training iterations to 100 for the manual data (ManualSet-100iter) suggesting that the poorer accuracy results from the limited amount of annotated data. Thus, the table demonstrates that using image stacks to automatically generate training data is an effective method and leads to impressive accuracy.



(a) Result of training with manually annotated data.



(b) Result of training with image stacks.



(c) 20× in-focus measure (d) 40× in-focus measure

Fig. 3. (a) Classification results on 20× and 40× stacks using classifiers trained with manually labeled data. The blue color represents in-focus assignment, and the red represents out-of-focus assignment. Ideally, a region is classified as in-focus near its focal plane, and as out-of-focus when far away. However, the lower-right regions of the 20× stack were always classified as out-of-focus, while almost all regions of the 40× stack were classified as in-focus. This indicates the poor generalization of the classifiers. (b) Classification results on the same 20× and 40× stacks as in (a) using classifiers trained with automatically generated data. The classification results show a clear trend throughout both stacks of each region going from out-of-focus to in-focus and back. (c) and (d) show plots of the percentage of in-focus tissue regions over each entire stack. The classification was carried out independently for each image of the stack, but they still show a clear and smooth trend of increasing towards the focal plane at $1\mu m$ as expected for the image stacks.

In Figure 2, we show some classification results overlaid on example test images, which were labeled by the pathologists as (a,b) "in-focus" and (c) "out-of-focus". The classifier indicated that the images were not fully in-focus or out-of-focus but rather included a mixture of in-focus and out-of-focus regions, and a closer inspection confirmed this. For example, the top-right portion of the tissue in Figure 2(c) (shown in blue) was in clear focus. The automatic classifier was sensitive enough to detect such labeling ambiguities.

To evaluate the generalization of the trained classifiers, we tested them on a 20× and a 40× image stack that were not used previously in either training or testing. Figure 3 shows these images scanned at different microscope depths overlaid with the classification results. Ideally, the in-focus/out-of-focus assignments by a classifier would change consistently along the stack, so that a region is classified as out-of-focus when scanned far away from the focal plane, and as in-focus when scanned at or close to the focal plane. However, as illustrated in Figure 3(a), the classifiers trained with the manual set generalized poorly in most of the regions. For example, it always classified the lower-right tissue regions of the 20× stack as out-of-focus and classified almost all regions of the 40× stack as in-focus. On the other hand, as shown in Figure 3(b), the classifiers trained with image stacks produced consistent results showing that most regions in both 20× and 40× stacks went from out-of-focus to in-focus and to out-of-focus again along the depth axis. In Figure 3(b), a small portion of tissues in the 20× scans was classified at out-of-focus throughout the entire stack. The pathologists verified that these are the edge regions of the tissue which lie at a different focal plane than the tissue and are indeed out-of-focus throughout the entire stack.

Figure 3(b) also shows that, in both stacks, as the stack images go from out-of-focus to in-focus, the in-focus regions propagate from the lower-right corner to the upper-left. This phenomenon seemed to indicate a problem that the stage which holds the tissue slide was not orthogonal to the optical axis of the camera. We reported this stage tilt problem that was automatically identified by the algorithms to the microscope design team, and they verified and corrected this issue. Thus, the classification results provided not only a means of feedback to the user but also crucial information for improving the microscope design.

Finally, we show the classification results for the WSI as the percentage of in-focus regions. Figures 3(c) and (d) present percentage measures for the stacks of Figure 3(b). Both curves show a smooth trend to increase towards the depth of $1\mu m$, indicating the true location of the focal plane, which was verified by a pathologist. Although the classification was done for each image independently, the trend follows closely the expected focus trend for image stacks.

5 Conclusions

We proposed an approach to evaluate the focus quality of whole slide images and introduced a way of generating training data from image stacks to train the classifiers. An extensive set of features were extracted and a boosting classifier was trained for classifying images as "in-focus" or "out-of-focus". Our experimental results not only showed that the classifiers trained with the image stacks

performed better than the limited manually labeled data, but also demonstrated the impressive accuracy of the results and their generalization ability.

References

1. Liron, Y., Paran, Y., Zatorsky, N.G., Geiger, B., Kam, Z.: Laser autofocusing system for high-resolution cell biological imaging. J. Microsc. 221, 145–151 (2006)
2. Santos, A., Ortiz de Solorzano, C., Vaquero, J.J., Pena, J.M., Malpica, N., del Pozo, F.: Evaluation of autofocus functions in molecular cytogenetic analysis. J. Microsc. 188, 264–272 (1997)
3. Firestone, L., Cook, K., Culp, K., Talsania, N., Preston, K.: Comparison of autofocus methods for automated microscopy. Cytometry 12, 195–206 (1991)
4. Liu, X.Y., Wang, W.H., Sun, Y.: Dynamic evaluation of autofocusing for automated microscopic analysis of blood smear and pap smear. J. Microsc 227, 15–23 (2007)
5. Yeo, T., Jayasooriah, S.H.O., Sinniah, R.: Autofocusing for tissue microscopy. Img. and Vis. Comp. 11, 629–639 (1993)
6. Nayar, S., Nakagawa, Y.: Shape from focus. IEEE PAMI 16(8), 824–831 (1994)
7. Dubois, F., Schockaert, C., Callens, N., Yourassowsky, C.: Focus plane detection criteria in digital holography microscopy by amplitude analysis. Opt. Express 14(13), 5895–5908 (2006)
8. Della Mea, V., Viel, F., Beltrami, C.A.: A pixel-based autofocusing technique for digital histologic and cytologic slides. Comput. Med. Imaging Graph 29, 333–341 (2005)
9. Yazdanfar, S., Kenny, K., Tasimi, K., Corwin, A., Dixon, E., Filkins, R.: Simple and robust image-based autofocusing for digital microscopy. Opt. Express 16, 8670–8677 (2008)
10. Sun, Y., Duthaler, S., Nelson, B.J.: Autofocusing in computer microscopy: selecting the optimal focus algorithm. Microsc. Res. Tech. 65, 139–149 (2004)
11. Padfield, D., Rittscher, J., Roysam, B.: Defocus and low CNR detection for cell tracking applications. In: MIAAB Workshop (2008)
12. Viola, P., Jones, M.: Robust real-time object detection. In: 2nd Int. Workshop on Statistical and Computational Theories of Vision Modeling, Learning, Computing and Sampling (July 2001)
13. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. Journal of Computer and System Sciences 55(1), 119–139 (1997)
14. Duda, R., Hart, P., Stork, D.: Pattern Classification. John Wiley & Sons, Chichester (2001)
15. Krotkov, E.: Focusing. IJCV 1(3), 223–237 (1988)
16. Subbarao, M., Choi, T., Nikzad, A.: Focus techniques. J. Opt. Eng. 32, 2824–2836 (1993)
17. Vollath, D.: Automatic focusing by correlative methods. J. Microsc 147, 279–288 (1987)
18. Huang, J., Mumford, D.: Statistics of Natural Images and Models. In: Proceedings IEEE Conference on Computer Vision and Pattern Recognition, pp. 541–547. IEEE Computer Society, Los Alamitos (1999)

Augmenting Capsule Endoscopy Diagnosis: A Similarity Learning Approach

S. Seshamani[1], R. Kumar[1], T. Dassopoulos[3], G. Mullin[2], and G. Hager[1]

[1] Department of Computer Science, Johns Hopkins University, Baltimore, MD
[2] Johns Hopkins Hospital, Baltimore, MD
[3] Washington University, School of Medicine St. Louis, MO*

Abstract. The current procedure for diagnosis of Crohn's disease (CD) from Capsule Endoscopy is a tedious manual process which requires the clinician to visually inspect large video sequences for matching and categorization of diseased areas (lesions). Automated methods for matching and classification can help improve this process by reducing diagnosis time and improving consistency of categorization. In this paper, we propose a novel SVM-based similarity learning method for distinguishing between correct and incorrect matches in Capsule Endoscopy (CE). We also show that this can be used in conjunction with a voting scheme to categorize lesion images. Results show that our methods outperform standard classifiers in discriminating similar from dissimilar lesion images, as well as in lesion categorization. We also show that our methods drastically reduce the complexity (training time) by requiring only one half of the data for training, without compromising the accuracy of the classifier.

1 Introduction

Wireless Capsule Endoscopy [1] is a non-invasive technology that is gaining popularity for diagnosis of Gastrointestinal (GI) diseases. The main advantage it provides, compared to traditional push endoscopy, is the increased comfort to the patient. The capsule endoscope, which is slightly larger than a common vitamin pill, contains lighting and imaging hardware, a wireless transmitter and a battery. It is taken orally by the patient and is propelled by peristalsis along the GI tract. Images are captured at the rate of 2fps by the device as it travels along the GI tract, and wirelessly transmitted to an archiving device that is attached to the patient's abdomen. The archiving device is then returned to the clinician who inspects the images (in the order of 50,000) to determine a diagnosis.

This diagnostic procedure, often 1-2 hours [1], requires the clinician to inspect several views of a pathology and evaluate it in a consistent manner. The low frame rate and lack of control of the imaging device, however, do not guarantee that subsequent images in the video sequence contain the same pathology.

* Supported in part by National Institutes of Health with Grant 5R21EB008227-02 and Johns Hopkins University internal funds.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 454–462, 2010.

Advances in the area of computer vision and machine learning can be used to improve both time and consistency of this evaluation process. In this paper, we consider the particular case of improving the diagnosis of CD. CD is an inflammatory bowel disease which is characterized by discrete, well-circumscribed erosions and ulcers (lesions) in the small intestine (Fig. 1). During inspection, the clinician locates lesions and assigns severity labels to them. Upon completing evaluation of each lesion individually, the overall diagnosis is determined.

There are two ways in which this process can be improved. The first is by automatic matching for locating multiple views of a selected pathology. Seshamani et al. [2] propose a meta matching procedure that incorporates several simple matchers and a binary decision function that determines whether a pair of images are similar or not. The second diagnostic improvement is the enhancement of CD lesion scoring consistency with the use of a predictor which can determine the severity of the lesion based on previously seen examples.

Both of these problems can be approached from a similarity learning perspective. Learning the decision function for meta matching is a similarity learning problem [3]. Lesion severity prediction is a multi-class classification problem which involves learning semantic classes of lesions based on appearance characteristics. Multi-class classification can also be approached from a similarity learning approach as shown in [3,4]. In this paper, we approach both of these problems as supervised pairwise similarity learning [5,6,7] problems .

2 Pairwise Similarity Learning

The pairwise similarity learning problem is the following: Given a pair of data points, determine if these two points are similar, based on previously seen examples of similar and dissimilar points. A function that performs this task is called a **pairwise similarity learner** (PSL). A PSL is made up of two parts: a representation function, and a classification function. In addition, the PSL is also required to be invariant to the ordering of pairs. One method of assuring order invariance is by imposing a symmetry constraint on the representation function [2]. However, doing so can introduce a loss of dimensionality and possibly a loss of information that may be relevant for the classification task.

Order invariance of the PSL can also be ensured by imposing symmetry constraints on the classifier. We refer to such a classification function as a **pairwise symmetric classifier**. Several SVM-based pairwise symmetric classifiers have been proposed in the literature [5,6,7]. Within the SVM framework, symmetry is imposed by ensuring that the kernel function satisfies order invariance. In all prior work concerning pairwise symmetric classifiers, a pair is described by only one type of feature and the underlying assumption is that one distance metric holds for the entire set of points. However, this assumption may not hold when multiple features are used to describe data. The area of Multiple Kernel Learning [8,9,10] has investigated several methods for combining features within the SVM framework.

In this paper, we present a novel pairwise similarity classifier for PSL using multiple features and nonsymmetric representations.

3 Mathematical Formulation

Consider a pair of images (I, J) and a set $\mathcal{X}$ consisting of m image descriptors (features). Applying any $X_i \in \mathcal{X}$ to each image in the pair generates a representation $\tilde{x} = (x_1, x_2)$ where $x_1 = \{X_i(I)\}$ and $x_2 = \{X_i(J)\}$. A label $y \in \{1, -1\}$ is associated with each pair $\tilde{x}$, where $y = 1$ implies a pair of similar images and $y = -1$ implies a pair of dissimilar images. The PSL problem can be written as follows: given a training set with n image pair representations and their associated labels $\mathcal{T}_m = \{(\tilde{x}_i, y_i) | i = 1 \dots n\}$, compute a classifier C that can predict the label of an unseen pair $\tilde{x}$:

$$C(\tilde{x}) = C((x_1, x_2)) = \begin{cases} 1, & \text{if } \tilde{x} \text{ represents a pair of similar images} \\ -1 & \text{otherwise} \end{cases} \tag{1}$$

Order invariance requires $C((x_1, x_2)) = C((x_2, x_1))$. We refer to this as the **pairwise symmetric constraint**. An SVM trained on the set $\mathcal{T}$ would classify an unseen pair $\tilde{x} = (x_1, x_2)$ as:

$$C(\tilde{x}) = \sum_{\forall (x_i, y_i) \in \mathcal{T}} \alpha_i y_i K(\tilde{x}, \tilde{x}_i) + b \tag{2}$$

where b and α_i's are learned from training examples and K is a Mercer kernel. It is easy to note that this classifier will satisfy the pairwise symmetric constraint if K satisfies: $K(\tilde{x}, \tilde{x}_i) = K((x_1, x_2), (x_{i1}, x_{i2})) = K((x_2, x_1), (x_{i1}, x_{i2}))$. We refer to such a kernel as a **pairwise symmetric kernel** (PSK).

3.1 PSKs for One Descriptor

Mercer Kernels can be generated from other Mercer Kernels by linear combinations (with positive weights) or elementwise multiplication [11]. We use this idea to generate PSKs from simpler Mercer Kernels. Let us assume that we have two pairs: (x_1, x_2) and (x_3, x_4) and a **base mercer kernel** K, which can operate on a pair of points. A PSK (which operates on two pairs of points) can be computed by symmetrization of the base kernel. In [5], a second order PSK called the **MLPK** is introduced: $\hat{K}((x_1, x_2), (x_3, x_4)) = (K(x_1, x_3) + K(x_2, x_4) - K(x_1, x_4) - K(x_2, x_3))^2$. This kernel is a linear combination of all second order combinations of the four base Mercer kernels. This kernel can be rewritten in terms of 3 PSKs as $\hat{K} = K_1 + 2K_2 - 2K_3$ where:

$$\begin{aligned} K_1 &= K(x_1, x_3)^2 + K(x_2, x_4)^2 + K(x_1, x_4)^2 + K(x_2, x_3)^2 \\ K_2 &= K(x_1, x_3)K(x_2, x_4) + K(x_1, x_4)K(x_2, x_3) \\ K_3 &= K(x_1, x_3)K(x_1, x_4) + K(x_1, x_3)K(x_2, x_3) + \\ &\quad K(x_2, x_4)K(x_1, x_4) + K(x_2, x_4)K(x_2, x_3) \end{aligned}$$

The MLPK kernel is different from a second order polynomial kernel due to the additional base kernels it uses. A classifier trained with the MLPK kernel is comparable to a classifier trained with a second order polynomial kernel on

double the amount of data (with pair orders reversed). SVM complexity can indeed be exponential in the number of training points (in the worst case) [12]. Secondly, a larger training dataset will generate more support vectors which increase run time complexity (classification time). Thus, the PSK is greatly beneficial in the reduction of both training and classification time.

3.2 PSKs with More than One Descriptor

For one descriptor, we obtain 3 second order PSKs(K_1, K_2 and K_3). So, given a set of m descriptors, we can generate a total of $3m$ second order PSKs: $Q = \{K'_i | i = 1 \dots 3m\}$. The problem now becomes the following: Given a set of PSKs find a weight vector $d \in \Re^{3m}$ that can generate a kernel $\hat{K} = \sum_i^{3m} d_i K'_i$ where $d_i \in d$, $K'_i \in Q$. We use Simple Multiple Kernel Learning (SimpleMKL) [8] for automatically learning these weights. This method initializes the weight vector uniformly and then performs a gradient descent on the SVM cost function to find an optimal weighting solution. The proposed **Generalized Pairwise Symmetric Learning** (GPSL) training algorithm is outlined below.
Input: Training set $\mathcal{T}_m$ and m base kernels.
Output: Weight Vector d_{best}, SVM parameters α and b

- For each of the m features, compute K_1,K_2 and K_3 (described in section 3.1) between all training pairs to generate the set $Q_{train} = \{K'_i | i = 1 \dots 3m\}$
- Apply SimpleMKL to find a weight vector d_{best}.
- Learn the SVM parameters α and b using a kernel generated as a linear combination of kernels in Q using d_{best}.

To predict similarity of an unseen pair $\tilde{x}$:

- Compute the set Q_{test} using the test point and training examples.
- Generate a linear combination of these kernels using d_{best}
- Predict the similarity of the pair using the learned α and b.

3.3 Multiclass Classification

The multiclass classification problem for images is as follows: Given a training set consisting of k images and their semantic labels $\mathcal{I} = \{(I_i, l_i) | i = 1 \dots k, l_i \in \{1 \dots p\}\}$, where I_is are the images and l_is are the labels belonging to one of p classes, compute a classifier that can predict the label of an unseen image I. From a similarity learning approach, this problem can be reformulated as a binary classification and voting problem: Given a training set of similar and dissimilar images, compute the semantic label of a new unseen image I. This requires two steps: 1) Learning similarities, and 2) Voting, to determine the label of an unseen image. We use the same method outlined in the GPSL algorithm above for similarity learning. Voting is then performed by selection of n voters from each semantic class who decide whether or not the new image is similar or dissimilar to themselves. We refer to this algorithm as **GPSL-Vote**:

- Given $\mathcal{I}$, compute a new training set consisting of all combinations of pairs and their similarity labels: $\mathcal{T} = \{((I_i, I_j)_k, y_k) | (I_i, l_i), (I_j, l_j) \in \mathcal{I}, y_k \in \{1, -1\}\}$ where $y_k = 1$ if $l_i = l_j$ and $y_k = -1$ otherwise .

– Train the GPSL using this set.

For a new image I,

– For each of the p semantic classes, select r representative images: $\{I_1 \dots I_r\}$ where (I_i, y_i) is such that $y_i = p$. This generates a set of $q = pr$ images.
– Compute a set of pairs by combining each representative image with the new image I: $\{(I, I_1) \dots (I, I_q)\}$
– Use the trained GPSL to predict which pairs are similar.
– For each semantic class, compute the number of similar pairs.
– Assign the new image I to the class with the maximum number of votes.

4 Experimental Results

4.1 Data Acquisition

Our Capsule Endoscopy study repository contains 47 anonymized studies acquired with an approved Institutional Review Board (IRB) protocol. Each study may contain up to 8 hours of data acquired at 2 images per second. The CE study database contains annotated images and videos containing Crohn's Disease (CD) lesions manually selected by our clinical collaborators. In all our experiments, the size of the selected image is on the order of 150X150 pixels. All our experiments are performed using MATLAB.

Annotated Lesion Data. Lesion annotation which was conducted by Dr. Dassopoulos involved selection of an image containing a lesion, and linking each image to a video segment consisting of 100-200 frames centered around the selected image. Each lesion image is then labelled as: Mild, Moderate or Severe (Fig. 1).

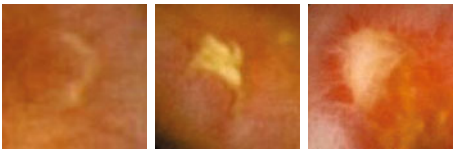


Fig. 1. Examples of Mild (left), Moderate (center) and Severe Crohn's Disease lesions

Annotated Matching Pairs. Each video sequence that is linked to a lesion image may contain multiple views of the same lesion. All views of the lesion within the video form a lesion set (Fig. 2). Various lesion sets contain between 2 and 25 image frames. We first manually segmented all regions containing lesions and then generated pairs within a lesion set as follows: Similar pairs were

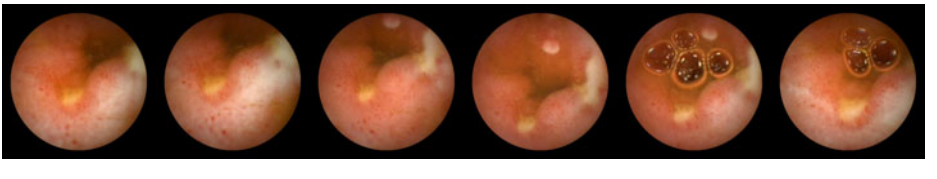


Fig. 2. A lesion set containing multiple views of the same lesion

generated by pairing segmented lesion regions within the lesion set. Dissimilar pairs were generated by pairing segmented lesion regions with 150X150 regions sampled 60 pixels away from the annotated center of a segmented lesion.

4.2 CE Lesion Matching Experiments

The objective of these experiments is to show the effectiveness of our pairwise symmetric classifier for discriminating between similar and dissimilar CE lesion images. Here, a lesion is considered similar to another one only if it is a different view of the same pathology.

Pair representation: Each image in a pair was represented by a set of descriptors: MPEG-7 Homogeneous Texture Descriptors (HTD) [13], color weighted histograms (WH) and patch intensities (PI). WHs were generated by dividing the color space into 11 bins and populating a feature vector with points weighted by their distance from the image center. PIs were generated by dividing the image into 16 patches and populating a vector with the mean intensity in each patch. The number of histogram bins and patches were determined empirically. A nonsymmetric pair consists of two sets of these descriptors stacked together. For the symmetric reprsentation, descriptors element-wise squared difference was carried out between the two sets. A **chi-squared base kernel** was used for WH and a **polynomial base kernel** of order 1 was used for the other two descriptors in all experiments in this section.

Experiments: We first show the following experimentally: MLPK with a non-symmetric representation is better than using a nonsymmetric kernel with a symmetric representation. Our full data set contains 724 lesion image pairs. We used 30% of the data for testing and the rest for training of the HTD, WH and PI descriptors. Training of the HTD, WH and PI with the symmetric representation resulted in accuracies of: 52.2%, 67.88% and 82% respectively. Training with the nonsymmetric representation resulted in

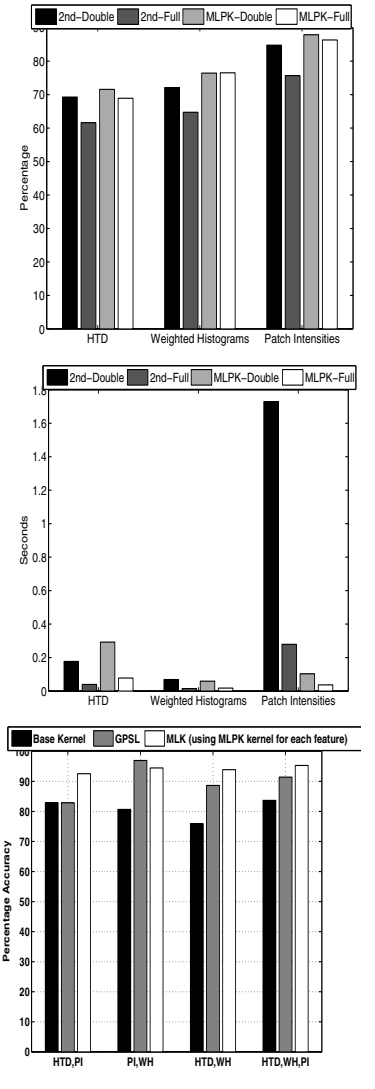


Fig. 3. Comparison of Accuracies (Top) and Training Time (Middle) for HTD, WH, PI using the full dataset and double dataset. Bottom: Comparison of the performance of the base kernel, MLPK and GPSL for various descriptor combinations

accuracies of: 64.2%, 75.22% and 88% for HTD,WH and PI respectively. Thus, we observe that generating a symmetric representation can cause a loss of information that is relevant for classification. Since MLPK is a second order kernel, the next question that arises is "How does an SVM trained with MLPK differ from an SVM trained with a second order polynomial kernel trained on double the data (generated by switching pair orders to enforce symmetry)?" To investigate this, we performed the following experiment. We first randomly selected 70% of the data as the **full training set** and the rest as the **full testing set**. We then doubled our data within each set by switching the order of pairs, to generate 1012 pairs for the **double training set** and 436 pairs for the **double testing set**. We performed classification (similar vs dissimilar pairs) using each descriptor 10 times by varying the type of training set and the type of kernel used (Table 1). Testing was always performed on the double testing set. Fig. 3 (bottom) shows the average accuracy and training time for all desriptors. The standard deviation of accuracies ranged from 2-4.5%. We observe that the MLPK greatly drops the training time when half the data is used without compromising much on accuracy, unlike the second order polynomial kernel.

Table 1. Four different experimental setups, varying in selection of a kernel and the amount of training data

Experiment	Kernel Type	Data
2nd-Double	2nd order poly	Double training set
MLPK-Double	MLPK	Double training set
2nd-Full	2nd order poly	Full training set
MLPK-Full	MLPK	Full training set

Finally, we show the effect of combining MLPKs generated from multiple features. We consider three algorithms for comparison: **SVM with a base kernel**, **SimpleMKL using MLPK** generated from the same base kernel (a total of m kernels) and **GPSL** (a total of 3m kernels also calculated from the same base kernel). We applied 5-fold CV to all three algorithms using all combinations of the three descriptors. Fig. 3 (bottom) shows the resulting accuracies. We observe that GPSL outperforms SVM with a base kernel in all cases. SimpleMLK with MLPK also performs much better than SVM with a base kernel in all cases, except the HTD descriptor.

4.3 CE Lesion Classification

Next, we considered the problem of classifying mild vs severe lesions (Fig. 1). A set of 10 lesions (5 in each class) were picked and three types of features were extracted: Haralick texture descriptor and Cross Correlation responses

Table 2. Accuracies of Different Classifiers used for lesion classification

Algorithm	Accuracy
SVM-Separate(Best Descriptor)	72 %
SVM-MKL	72 %
GPSL (Similarity learner)	71.75 %
GPSL-Vote	76 %

of the blue and green bands with the same bands of a template lesion image. We then performed and compared three classification experiments: SVM with each descriptor separately (SVM-Separate) to directly classify lesion images, SVM with all features combined by SimpleMKL (SVM-MKL) to directly classify lesion images and finally with GPSL-Vote (which uses pairwise similarity learning). CV in all cases was performed on a "leave-two-out" basis, where the testing set was made up of one image from each class. All other images formed the training set. This generated a total of 25 total runs. In the case of GPSL-Vote, the similarity training dataset was generated using all combinations of pairs which are in the training set (totally 45 pairs). All pairs containing the two left-out images and a training image formed the test set. The results shown in Table 2 show the mean accuracy of all 25 runs. We observe that the SVM-MKL algorithm does only as well as the best classifier. However, GPSL-vote outperforms this, even for a small dataset with a small number of features.

5 Conclusion

We have demonstrated the use of pairwise symmetric learners for lesion matching and categorization. Our results show that the use of pairwise symmetric classifiers not only allows for nonsymmetric representation of data but also drastically reduces training time without compromising the accuracy of the classifier. We have also demonstrated the use of PSK's across multiple features for lesion match classification as well as semantic lesion classification in CE imagery. One extension of this work is the application of our classifier to a real-time matching system to increase the time efficiency of the clinician. Although the GPSL-vote experimentation was performed on a small dataset, the results are very encouraging. Some of the natural extensions of this work include experimentation with more descriptors and larger datasets for GPSL-vote.

References

1. Iddan, G., Meron, G., Glukhovsky, A., Swain, P.: Wireless capsule endoscopy. Nature 405(6785), 417 (2000)
2. Seshamani, S., Rajan, P., Kumar, R., Girgis, H., Mullin, G., Dassopoulos, T., Hager, G.: A meta registration framework for lesion matching. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 582–589. Springer, Heidelberg (2009)
3. Chen, Y., Garcia, E.K., Gupta, M.R., Rahimi, A., Cazzanti, L.: Similarity-based classification: Concepts and algorithms. JMLR 10, 747–776 (2009)
4. Cazzanti, L., Gupta, M.R.: Local similarity discriminant analysis. In: ICML (2007)
5. Vert, J.P., Qiu, J., Noble, W.S.: A new pairwise kernel for biological network inference with support vector machines. BMC Bioinformatics 8(S-10) (2007)
6. Kashima, H., Oyama, S., Yamanishi, Y., Tsuda, K.: On pairwise kernels: An efficient alternative and generalization analysis. In: Theeramunkong, T., Kijsirikul, B., Cercone, N., Ho, T.-B. (eds.) PAKDD 2009. LNCS, vol. 5476, pp. 1030–1037. Springer, Heidelberg (2009)

7. Oyama, S., Manning, C.D.: Using feature conjunctions across examples for learning pairwise classifiers. In: Boulicaut, J.-F., Esposito, F., Giannotti, F., Pedreschi, D. (eds.) ECML 2004. LNCS (LNAI), vol. 3201, pp. 322–333. Springer, Heidelberg (2004)
8. Rakotomamonjy, A., Bach, F.R., Canu, S., Grandvalet, Y.: Simplemkl. JMLR 9 (2008)
9. Varma, M., Babu, B.R.: More generality in efficient multiple kernel learning. In: ICML, pp. 1065–1072 (June 2009)
10. Gehler, P., Nowozin, S.: Let the kernel figure it out: Principled learning of pre-processing for kernel classifiers. In: CVPR (2009)
11. Cristianini, N., Shawe-Taylor, J.: An Introduction to Support Vector Machines: and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge (2000)
12. Gärtner, B., Giesen, J., Jaggi, M.: An exponential lower bound on the complexity of regularization paths. In: CoRR (2009)
13. Manjunath, B., Ohm, J., Vasudevan, V., Yamada, A.: Color and texture descriptors. IEEE CSVT 11(6), 703–715 (2001)

A Fully Automated Approach to Segmentation of Irregularly Shaped Cellular Structures in EM Images

Aurélien Lucchi*, Kevin Smith, Radhakrishna Achanta,
Vincent Lepetit, and Pascal Fua

Computer Vision Lab, Ecole Polytechnique Fédérale de Lausanne, Switzerland

Abstract. While there has been substantial progress in segmenting natural images, state-of-the-art methods that perform well in such tasks unfortunately tend to underperform when confronted with the different challenges posed by electron microscope (EM) data. For example, in EM imagery of neural tissue, numerous cells and subcellular structures appear within a single image, they exhibit irregular shapes that cannot be easily modeled by standard techniques, and confusing textures clutter the background. We propose a fully automated approach that handles these challenges by using sophisticated cues that capture global shape and texture information, and by learning the specific appearance of object boundaries. We demonstrate that our approach significantly outperforms state-of-the-art techniques and closely matches the performance of human annotators.

1 Introduction

State-of-the-art segmentation algorithms which perform well on standard natural image benchmarks such as the Pascal VOC dataset [7] tend to perform poorly when applied to EM imagery. This is because the image cues they rely upon tend not to be discriminative enough for segmenting structures such as mitochondria. As shown in Fig. 1(a), they exhibit irregular shapes not easily captured using standard shape modeling methods. Their texture can easily be confused with that of groups of vesicles or endoplasmic reticula. Mitochondrial boundaries are difficult to distinguish from other membranes that share a similar appearance. Overcoming these difficulties requires taking all visible image cues into account simultaneously. However, most state-of-the-art techniques are limited in this respect. For example, TextonBoost uses sophisticated texture and boundary cues, but simple haar-like rectangular features capture shape [14]. In [5], SIFT descriptors capture local texture and gradient information, but shape is ignored.

Previous attempts at segmenting neural EM imagery include a normalized cuts based approach in [8]. More recently, [3] used a level set approach which is sensitive to initialization and limited to one object. [15] is an active contour approach designed to detect elliptical blobs but fails to segment mitochondria which often take non-ellipsoid shapes. In [4], a convolutional neural network considers only local information using a watershed-based supervoxel segmentation. Finally, [12] uses a classifier on texton features to learn mitochondrial texture, but ignores shape information.

* This work was supported in part by the MicroNano ERC project and by the Swiss National Science Foundation Sinergia Project CRSII3-127456.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 463–471, 2010.

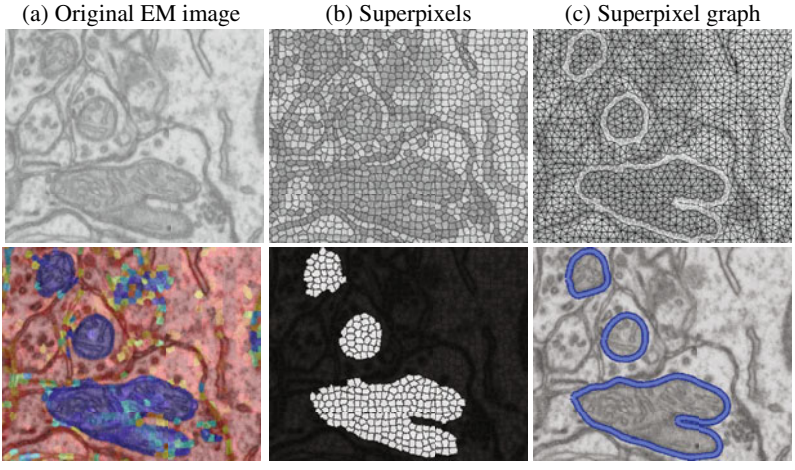


Fig. 1. Overview. **(a)** A detail of the original EM image. **(b)** Superpixel over-segmentation. **(c)** Graph defined over superpixels. White edges indicate pairs of superpixels used to train an SVM that predicts mitochondrial boundaries. **(d)** SVM prediction where blue indicates a probable mitochondrion. **(e)** Graph cut segmentation. **(f)** Final results after automated post-processing. Note: the same image is used in this figure for clarity; images in the training & testing sets are disjoint.

In this paper, we propose to overcome these limitations by:

1. **Using all available image cues simultaneously:** We consider powerful shape cues that do not require an explicit shape model in addition to texture and boundary cues.
2. **Learning the appearance of boundaries on a superpixel graph:** We train a classifier to predict where mitochondrial boundaries occur using these cues.

An overview of our approach appears in Fig. 1. We first produce a superpixel over-segmentation of the image to reduce computational cost and enforce local consistency. The superpixels define nodes in a graph used for segmentation. We then extract sophisticated shape, texture, and boundary cues captured by Ray [10] and Rotational [9] features for each superpixel. Support vector machine (SVM) classifiers are trained on these features to recognize the appearance of superpixels belonging to mitochondria, as well as pairs of superpixels containing a mitochondrial membrane. Classification results are converted to probabilities, which are used in the *unary* and *pairwise* terms of a graph-cut algorithm that segments the superpixel graph. Finally, an automated post-processing step smooths the segmentation. We show qualitatively and quantitatively that our approach yields substantial improvements over existing methods. Furthermore, whatever mistakes remain can be interactively corrected using well known methods [6].

2 Our Approach

2.1 Superpixel Over-Segmentation

Our first step is to apply a novel *k-means* based algorithm [13] to aggregate nearby pixels into *superpixels* of nearly uniform size whose boundaries closely match true image

boundaries, as seen in Fig. 1(b). It has been shown that using superpixels can be advantageous because they preserve natural image boundaries while capturing redundancy in the data [5]. Furthermore, superpixels provide a convenient primitive from which to compute local image features while reducing the complexity of the optimization by reducing the number of nodes in the graph.

2.2 Segmentation by Graph Partitioning

Graph-cuts is a popular approach to segmentation that splits an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ into partitions by minimizing an objective function [16]. As shown in Fig. 1(c), the graph nodes $\mathcal{V}$ correspond to superpixels x_i. Edges $\mathcal{E}$ connect neighboring superpixels. The objective function takes the form

$$E(c|x, w) = \sum_i \underbrace{\psi(c_i|x_i)}_{\text{unary term}} + w \sum_{(i,j)\in\mathcal{E}} \underbrace{\phi(c_i, c_j|x_i, x_j)}_{\text{pairwise term}} , \tag{1}$$

where $c_i \in \{foreground, background\}$ is a class label assigned to superpixel x_i. The so-called *unary* term ψ assigns to each superpixel its potential to be foreground or background based on a probability $P(c_i|\mathbf{f}(x_i))$ computed from the output of an SVM

$$\psi(c_i|x_i) = \frac{1}{1 + P(c_i|\mathbf{f}(x_i))}. \tag{2}$$

The *pairwise* term ϕ assigns to each pair of superpixels a potential to have similar or differing labels (indicating boundaries), based on a second SVM output

$$\phi(c_i, c_j|x_i, x_j) = \begin{cases} \frac{1}{1+P(c_i,c_j|\mathbf{f}(x_i),\mathbf{f}(x_j))} & \text{if } c_i \neq c_j, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The weight w in Eq. 1 controls the relative importance of the two terms. Our segmentation is achieved by minimizing Eq. 1 using a mincut-maxflow algorithm.

2.3 Superpixel-Based Shape and Local Features

The SVMs in Eqs. 2 and 3 predict which superpixels contain mitochondria and which neighboring superpixels contain a mitochondrial boundary. As discussed in Section 1, shape, texture, and boundary cues are all essential to this process. Features $\mathbf{f}(x_i)$ extracted from the image at superpixel x_i combine these essential cues

$$\mathbf{f}(x_i) = [\mathbf{f}^{\text{Ray}}(x_i)^\top, \mathbf{f}^{\text{Rot}}(x_i)^\top, \mathbf{f}^{\text{Hist}}(x_i)^\top]^\top , \tag{4}$$

where $\mathbf{f}^{\text{Ray}}$ represents Ray descriptors that capture object shape, $\mathbf{f}^{\text{Rot}}$ are rotational features describing texture and boundaries [9], and $\mathbf{f}^{\text{Hist}}$ are histograms describing the local intensity. These features, shown in Fig. 2, are detailed below.

Ray Descriptors describe the shape of local objects for each point in the image in a way that standard shape modeling techniques can not. Typically, other methods represent object shape using contour templates [1] or fragment codebooks [2]. While these

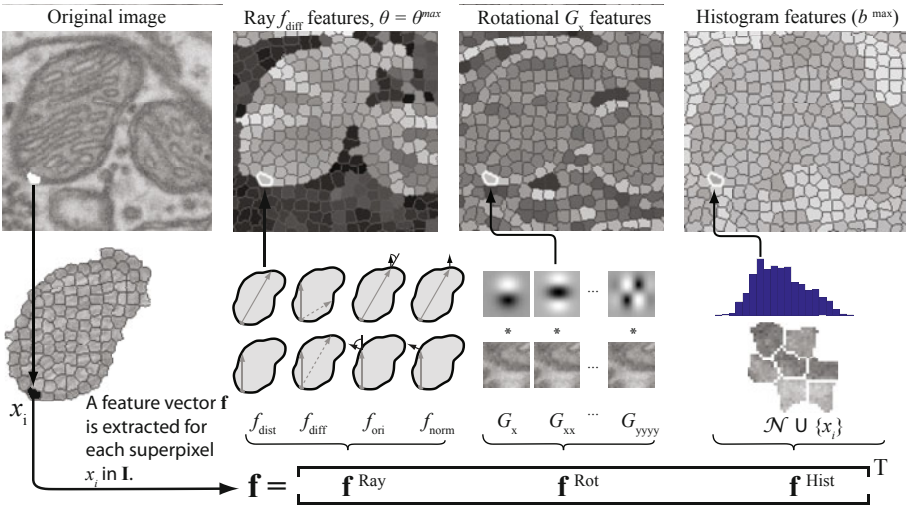


Fig. 2. For each superpixel, the SVM classifiers in Eqs. 2 and 3 predict the presence of mitochondria based on a feature vector $\mathbf{f}$ we extract. $\mathbf{f}$ captures shape cues with a Ray descriptor $\mathbf{f}^{\text{Ray}}$, texture and boundary cues with rotational features $\mathbf{f}^{\text{Rot}}$, and intensity cues in $\mathbf{f}^{\text{Hist}}$.

approaches can successfully segment a single object with known shape, they tend to fail when the shape is highly variable or when many objects appear in the image.

For a given point x_i in the image, four types of Ray features are extracted by projecting rays from x_i at regular angles $\mathbf{\Theta} = \{\theta_1, \ldots, \theta_N\}$ and stopping when they intersect a detected edge (r) [10]. The distance from x_i to r form the first type of feature f_{dist}. The other three types of features compare the relative distance from x_i to r for rays in two different directions (f_{diff}), measure the gradient strength at r (f_{norm}), and measure the gradient orientation at r relative to the ray (f_{ori}). While [10] uses individual Ray features as AdaBoost learners, we aggregate all features extracted for a single point into a *Ray descriptor* $\mathbf{f}^{\text{Ray}} = [f_{\text{dist}}\ f_{\text{diff}}\ f_{\text{norm}}\ f_{\text{ori}}]^\top$. We make it rotation invariant by shifting the descriptor elements so that the first element corresponds to the longest ray. Fig. 3 demonstrates the Ray descriptor's ability to compactly represent object shape.

Rotational Features capture texture and image cues indicating boundaries such as edges, ridges, crossings and junctions [9]. They are projections of image patches around a superpixel center x_i into the space of Gaussian derivatives at various scales, rotated to a local orientation estimation for rotational invariance.

Histograms complement $\mathbf{f}^{\text{Ray}}$ and $\mathbf{f}^{\text{Rot}}$ with simple intensity cues from superpixel x_i's neighborhood $\mathcal{N}$. $\mathbf{f}^{\text{Hist}}$ is written $\mathbf{f}^{\text{Hist}}(\mathbf{I}, x_i) = \sum_{j \in \mathcal{N} \cup \{i\}} h(\mathbf{I}, x_j, b)$ where $h(\mathbf{I}, x_j, b)$ is a b-bin histogram extracted from $\mathbf{I}$ over the pixels contained in superpixel x_j.

2.4 Learning Object Boundaries

Most graph-cut approaches model object boundaries using a simple pairwise term

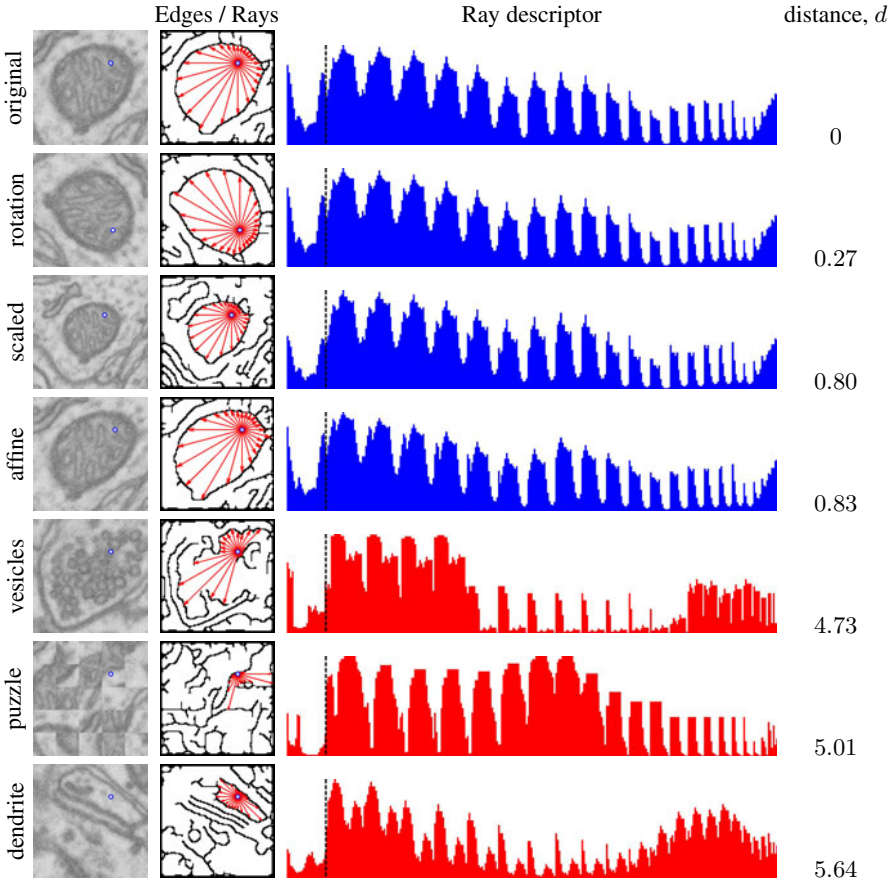


Fig. 3. Ray descriptors built from features in [10] provide a compact representation of local shape for each point in an image. The descriptors are stable when subjected to rotation, scale, and affine transformations, but change dramatically for other shapes including vesicles, dendrites, and randomly rearranged tiles from the original image (puzzle). d is the Euclidean distance between the descriptor extracted from the original image and descriptors extracted from other images.

$$\phi(c_i, c_j | x_i, x_j) = \begin{cases} \exp\left(-\frac{||I(x_i)-I(x_j)||^2}{2\sigma^2}\right), & \text{if } c_i \neq c_j \\ 0 & \text{, otherwise,} \end{cases} \tag{5}$$

which favors cuts at locations where color or intensity changes abruptly, as in [16]. While similar expressions based on Laplacian zero-crossings and gradient orientations exist [16], very few works go beyond this standard definition. As illustrated in Fig. 4 *(left)*, this approach results in a poor prediction of where mitochondrial boundaries actually occur, as strong gradients from other membranes cause confusion. By learning what image characteristics indicate a true object boundary, we can improve the segmentation [11]. We train an SVM using features extracted from pairs of

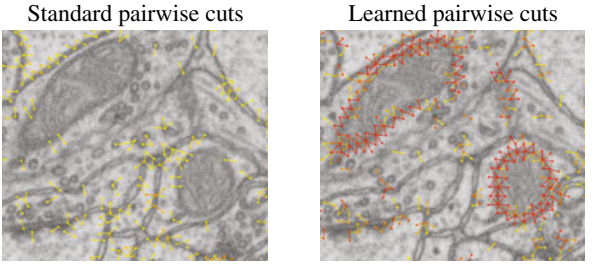


Fig. 4. *(left)* Boundaries predicted by a standard pairwise term (Eq. 5) correspond to strong gradients, but not necessarily to mitochondrial boundaries. *(right)* A learned pairwise term (Eq. 3) using more sophisticated cues $[\mathbf{f}_i^\top,\ \mathbf{f}_j^\top]^\top$ results in better boundary predictions. Red lines indicate strong probable boundaries, yellow lines indicate weaker boundaries.

superpixels containing true object boundaries, indicated by white graph edges in Fig. 1. The pairwise feature vector $\mathbf{f}_{i,j}$ is a concatenation of $\mathbf{f}_i$ and $\mathbf{f}_j$ extracted from each superpixel $\mathbf{f}_{i,j} = [\mathbf{f}_i^\top,\ \mathbf{f}_j^\top]^\top$, providing rich image cues for the SVM to consider.

3 Results

We tested our approach on a data set consisting of 23 annotated high resolution EM images. Each image is 2048×1536 pixels, and the entire data set contains 1023 total mitochondria. We used $k = 5$ k-fold cross validation for training and testing. Our evaluation compares segmentation results for the following methods:

TextonBoost A boosted texton-based segmentation algorithm [14],
Fulkerson09 A superpixel-based algorithm using SIFT features [5],
Standard-f* Our algorithm trained with the *standard* pairwise term of Eq. 5 and histogram and rotational features $[\mathbf{f}^{\text{Hist}\top}\mathbf{f}^{\text{Rot}\top}]^\top$,
Standard-f Our algorithm trained with the *standard* pairwise term of Eq. 5 and feature vector $\mathbf{f}$ incorporating shape and texture cues given in Eq. 4.
Learned-f *Our complete algorithm* trained with the *learned* pairwise term of Eq. 3 and feature vector $\mathbf{f}$ incorporating shape and texture cues given in Eq. 4.

Parameter settings for [14] used 50 textons and 2000 rounds of boosting. For [5], Quickshift superpixels were used, and SIFT descriptors were extracted over 9 scales at a fixed orientation and quantized into 50 clusters. For our approach, we used superpixels containing approximately 100 pixels, extracted Rays at 30° angles, computed rotational

Table 1. Segmentation Results

	TextonBoost [14]	Fulkerson09 [5]	Standard-$\mathbf{f}^*$	Standard-$\mathbf{f}$	Learned-$\mathbf{f}$
Accuracy	95%	96%	94%	96%	**98%**
VOC score [7]	61%	69%	60%	68%	**82%**

features using first to fifth Gaussian derivatives with $\sigma = \{3, 6, 9, 12\}$, and built histograms with $b = 20$ bins. A post-processing step depicted in Fig. 1(f) was used to smooth the results produced by all the algorithms.

Discussion. Table 1 summarizes results for the entire data set. Our approach achieved a pixel-wise accuracy of 98%. By the same metric, TextonBoost and Fulkerson09 also performed well, but visually the results are inferior, as seen in Fig. 5. This is because mitochondria account for very few pixels in the image. The VOC score $= \frac{TP}{TP+FP+FN}$, introduced in [7][1], is designed to be more informative in such cases, and reflects the superior quality of our segmentations. Because it is pixel-based and lacks shape cues, TextonBoost poorly estimates mitochondrial membranes. The use of superpixels in Fulkerson09 seems to improve results slightly over [14], but the lack of shape cues or learned boundaries still degrades its performance. Comparing the Standard-$\mathbf{f}^*$ and Standard-$\mathbf{f}$ variations of our approach, we see that adding shape cues boosts performance, and learning boundaries in the pairwise term leads to a further increase in Learned-$\mathbf{f}$.

4 Conclusion

We proposed a fully automated approach to segment irregularly shaped cellular structures that outperforms state-of-the-art algorithms on EM imagery. We also demonstrated that Ray descriptors increase performance by capturing shape cues without having to define an explicit model. Finally, we showed that a learning approach to the pairwise term of the energy function further helps find true object boundaries.

Acknowledgements. We wish to thank Graham Knott and Marco Cantoni for providing us with high-resolution imagery and invaluable advice. We also thank German Gonzalez for providing code for Rotational Features.

References

1. Ali, A., Farag, A., El-Baz, A.: Graph Cuts Framework for Kidney Segmentation With Prior Shape Constraints. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 384–392. Springer, Heidelberg (2007)
2. Levin, A., Weiss, Y.: Learning to Combine Bottom-Up and Top-Down Segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 581–594. Springer, Heidelberg (2006)
3. Vazquez-Reina, A., Miller, E., Pfister, H.: Multiphase Geometric Couplings for the Segmentation of Neural Processes. In: CVPR (2009)
4. Andres, B., Koethe, U., Helmstaedter, M., Denk, W., Hamprecht, F.: Segmentation of Sbfsem Volume Data of Neural Tissue by Hierarchical Classification. In: Rigoll, G. (ed.) DAGM 2008. LNCS, vol. 5096, pp. 142–152. Springer, Heidelberg (2008)
5. Fulkerson, B., Vedaldi, A., Soatto, S.: Class Segmentation and Object Localization With Superpixel Neighborhoods. In: ICCV (2009)
6. Rother, C., Kolmogorov, V., Blake, A.: "GrabCut" - Interactive Foreground Extraction Using Iterated Graph Cuts. In: SIGGRAPH (2004)

[1] TP=true positives, FP=false positives, FN=false negatives.

7. Everingham, M., Van Gool, L., Williams, C.K.I., Winn, J., Zisserman, A.: The PASCAL Visual Object Classes Challenge (VOC 2010) Results (2010)
8. Frangakis, A., Hegerl, R.: Segmentation of two- and three-dimensional data from electron microscopy using eigenvector analysis. Journal of Structural Biology (2002)
9. Gonzalez, G., Fleuret, F., Fua, P.: Learning Rotational Features for Filament Detection. In: Conference on Computer Vision and Pattern Recognition (June 2009)
10. Smith, K., Carleton, A., Lepetit, V.: Fast Ray Features for Learning Irregular Shapes. In: ICCV (2009)
11. Prosad, M., Zisserman, A., Fitzgibbon, A., Kumar, M., Torr, P.: Learning Class-Specific Edges for Object Detection and Segmentation. In: Kalra, P.K., Peleg, S. (eds.) ICVGIP 2006. LNCS, vol. 4338, pp. 94–105. Springer, Heidelberg (2006)
12. Narashimha, R., Ouyang, H., Gray, A., McLaughlin, S., Subramaniam, S.: Automatic Joint Classification and Segmentation of Whole Cell 3D Images. Pattern Recognition 42(2007), 1067–1079 (2009)
13. Radhakrishna, A., Shaji, A., Smith, K., Lucchi, A., Fua, P., Susstrunk, S.: SLIC Superpixels. Technical Report 149300, EPFL (June 2010)
14. Shotton, J., Winn, J., Rother, C., Criminisi, A.: Textonboost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954. Springer, Heidelberg (2006)
15. Thévenaz, P., Delgado-Gonzalo, R., Unser, M.: The Ovuscule. PAMI (to appear, 2010)
16. Boykov, Y., Jolly, M.: Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images. In: ICCV (2001)

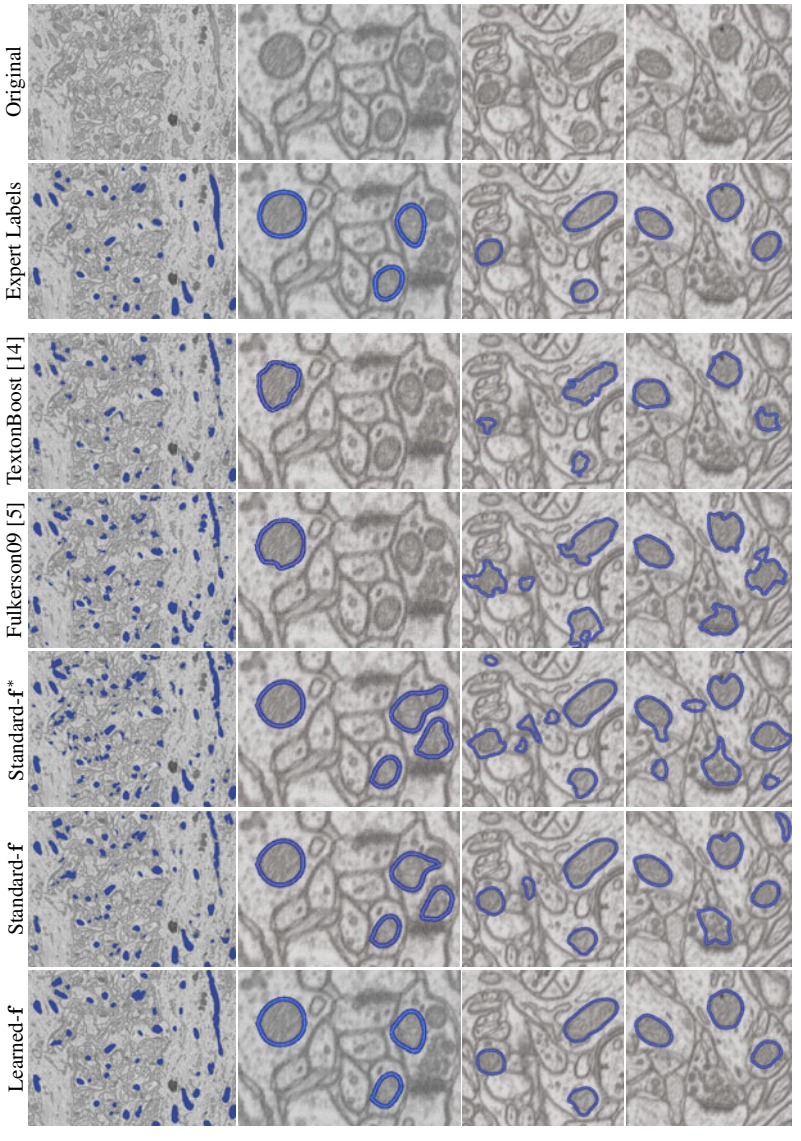


Fig. 5. Segmentation results on EM images. Column 1 contains the 2048×1536 micrograph at reduced resolution. Columns 2-4 contain details from column 1. Row 1 contains the original EM image. Row2 contains the expert annotations. Further rows contain results of the various methods. The lack of shape cues and learned boundaries result in inaccurate segmentations for TextonBoost and Fulkerson09, especially near distracting textures and membranes. Our method without shape or learned boundaries, Standard-$\mathbf{f}^*$, performs similarly. By injecting shape cues in Standard-$\mathbf{f}$, we see a significant improvement as more mitochondria-like shapes appear in the segmentation. However, some mistakes in the boundary persist. In Learned-$\mathbf{f}$ we add the learned pairwise term, eliminating the remaining errors and producing a segmentation that very closely resembles the human annotation.

Automatic Neuron Tracing in Volumetric Microscopy Images with Anisotropic Path Searching

Jun Xie, Ting Zhao, Tzumin Lee, Eugene Myers*, and Hanchuan Peng*

Janelia Farm Research Campus,
Howard Hughes Medical Institute, Ashburn, Virginia USA

Abstract. Full reconstruction of neuron morphology is of fundamental interest for the analysis and understanding of neuron function. We have developed a novel method capable of tracing neurons in three-dimensional microscopy data automatically. In contrast to template-based methods, the proposed approach makes no assumptions on the shape or appearance of neuron's body. Instead, an efficient seeding approach is applied to find significant pixels almost certainly within complex neuronal structures and the tracing problem is solved by computing an graph tree structure connecting these seeds. In addition, an automated neuron comparison method is introduced for performance evaluation and structure analysis. The proposed algorithm is computationally efficient. Experiments on different types of data show promising results.

1 Introduction

The tasks of extracting neuron structures from microscope images and accurately measuring their topologic properties are important for a large number of studies in biology and medicine. Although this is an extensively studied problem in the literature and there are some commercial products available, no automated method has promised reconstruction that requires no or few manual corrections. This prevents the implementation of high-thoroughput neuron structure analysis, without which biologists can be easily overwhelmed by their data.

Most existing methods are based upon the fact that a neuron is a branching tree structure. They formulate the problem of reconstructing a neuron as determining where the branches go and how they bifurcate. For this reason, the problem is also called neuron tracing. The branches can extend in any direction in the three-dimensional (3D) space. Therefore a successful strategy for realistic tracing applications has to operate in 3D. In 1994, Cohen *et al.* [1] introduced a 3D tracing method based on skeletonization and graph extraction. The same strategy was applied by Koh *et al.* [2] but with a modified medial axis to locate dendritic branch centerlines. However, 3D skeletonlization is computationally expensive and the noise in the microscopy data presents a significant challenge

* Equal senior authors.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 472–479, 2010.

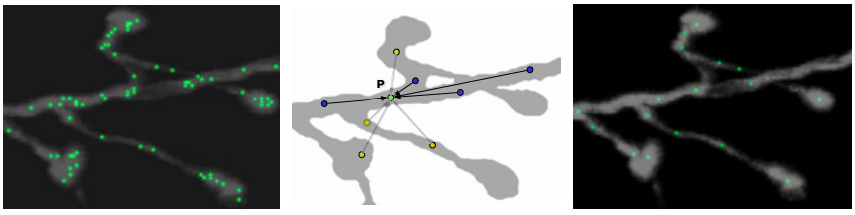


Fig. 1. An example of seeding. (Left) Initial seed candidates (green) from local maximum. (Middle) The diagram showing seed refinement. The ray between one candidate to candidate p determines whether it is an valid voter (blue) or not (yellow). (Right) The final seeds for tracing.

for the thinning operation. Based on structure continuity, parametric models were proposed for filament tracing [3]. Alkofahi *et al.* [4] proposed applying predetermined deformable templates to detect the local dendritic boundary and then performing an exploratory tracing of the dendrite. He *et al.* [5] proposed combining local color-based estimates of the probability that an image pixel belongs to a filament with the global tree properties of the complete set of filaments. If the neurite structures in the stack are all connected, template based tracing works well. However, for common realistic cases, the structures of interest appear to be discontinuous due to imaging conditions and noise. An optimization scheme based on the Minimum Spanning Tree (MST) [6] is usually applied to tackle this challenge.

2 Method

In this work, we follow the optimal tree idea to reconstruct the 3D topology of neuronal structures. Similar to the human operator's exercise, our method first detects a set of seeds which should be dense enough to describe the major characteristics of the neuronal structures, but also be considerably compact for redundancy control and computational efficiency.

Optimal Seeding. The ideal set of seeds consists of the points on the underlying objects which stands for a good representative to each neurite segment. Before detecting seed points, a global thresholding on the input 3D data is applied to remove noise (isolated particles) and save computation time. Based on the fact that noise often appears as small local maximal regions in the background area of microscope images, the global threshold is obtained by applying the triangle-thresholding method [7] on the histogram of local intensity maxima in the input image.

Our seeding method consists of two steps. In the first step, candidate seeds are detected by searching local intensity maixma. The local window is a cube on the stack grid and only one candidate is identified in each valid cube. A cube is considered as valid only if it contains sufficient foreground voxels which can be measured with the signal-to-noise ratio $\mathrm{SNR}_c = (\mu_f - \mu_b)/\sigma_b$, where μ_f is the mean intensity value of the foreground in the cube, μ_b and σ_b denotes

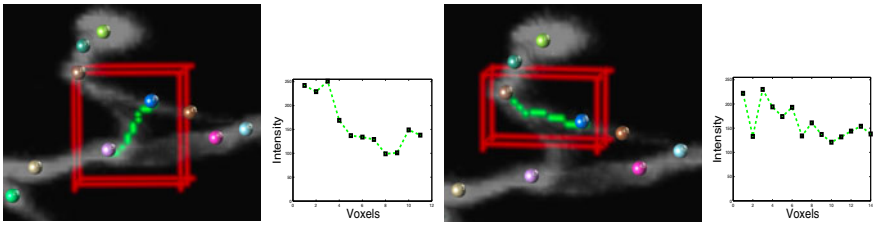


Fig. 2. An example of path searching. (Left) The strongest path detected for seed #37 (blue) with the isotropic weight function in Eq. (1). The corresponding intensity profile of the path is shown beside. (Right) The strongest path detected with the anisotropic weight function in Eq. (2) which favors smoothness.

the background's mean and standard deviation, respectively. In order to remove the high-frequency components and improve the detection accuracy, a Gaussian kernel is applied to filter the local intensity maxima. An example of this initial sampling is presented in Fig. 1(Left).

Secondly, a decimation operation is applied to those candidates for final reliable seeds. The local scale of the neurite structure is first estimated with the distance transformation [8]. Then for candidates with lower scale values, indicating they are isolated speckles or close to boundary, votes from other candidates are collected for refinement decision. In specific, for candidate p which has the lowest scale in the set, other candidates can recommend for elimination if they have a strong ray connection R to p ($\bar{I}_R > I_t$) (Fig. 1(Middle)). Candidate p will be removed from the set if there are recommendations from diverse orientations (*e.g.*, the maximum included angle of the voter rays is considerably large). This condition lets us avoid eliminating terminal seeds mistakenly. This seeding approach is robust to the intensity variation and leads to varying seeding density (Fig. 1(Right)) according to the local complexity of the structures, thus simplifying the tracing problem.

Linking Strength among Seeds. With the representative seeds detected above, tracing neuronal structures can be considered as finding the optimal connections of those seeds. Considering the image grid as a discretized weighted grid graph $G = (V, E, W)$, where vertex set V consists of the grid points, E contains the regular straight line segments in the grid and W are some features associated with the segments. By modeling the stack grid as a graph, one voxel is connected with its 26-connected neighbors. Given a pair of nodes v_0 to v_n, a path between them is a list of unique nodes. The cost of a path is the sum of each edge weight in the path.

Let $\mathcal{P}$ denote the set of all unit-speed paths in the search region $\mathcal{R}$ from v_0 to v_n, *i.e.*, the set of functions $\ell : [0, N] \to \mathcal{R}$, for which $\ell(0) = v_0$ and $\ell(N) = v_n$. The formulation of finding a shortest path between two vertices is to compute a path $\ell^* \in \mathcal{P}$ such that $\mathcal{W}(\ell^*) = \min_{\ell \in \mathcal{P}} \mathcal{W}(\ell)$, where $\mathcal{W} : \mathcal{P} \to [0, \infty]$ denotes the overall cost of the path: $\mathcal{W}(\ell) = \sum_{t=0}^{N} \rho(\ell(t))$, for each ℓ in $\mathcal{P}$. One of the efficient algorithm to find shortest path on the sparse graph is the one introduced

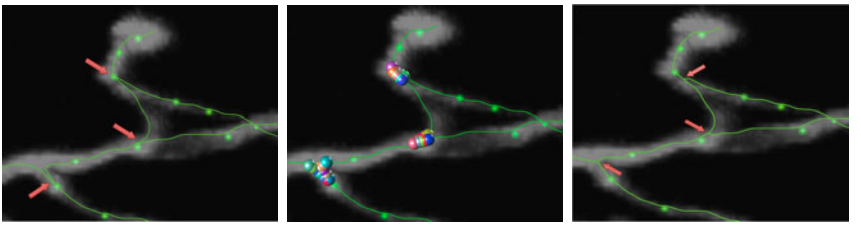


Fig. 3. Path Merging. (Left) Initial paths among seeds. (Middle) The 'touching' positions between nearby paths. (Right) The updated paths with the new branching locations (red arrows).

by Johnson [9] by storing the path lengths with a priority queue. The complexity of Johnson's algorithm is $O(|E|\log|V|)$.

The cost weighting function ρ varies according to the voxel's relative positions $\rho(v_i) \propto d(v_i)$ where $d(v_i) = |v_i - v_{i+1}|$. To emphasize the difference between foreground and background regions, a region-based cost term $\mathcal{I}$ is incorporated in the weighting function as:

$$\mathcal{W}(\ell) = \sum_{i=0}^{N-1} d(v_i)\mathcal{I}(v_i) = \sum_{i=0}^{N-1} d(v_i)\frac{\mathcal{S}(v_i) + \mathcal{S}(v_{i+1})}{2}, \tag{1}$$

where $\mathcal{S}(\cdot)$ is a Sigmoid-filtered version of the input stack which generates a large weight value if two voxels belong to different regions, indicating the path is crossing a significant boundary.

One limitation of the above cost-weighting function is that it considers only situations where the metrics are isotropic. It cannot account for the inherent properties of neurite structures and doesn't provide a smoothing constraint for the path searching. In our case, the neurite structures have a tree-liked topology and most of the components are of considerable smoothness. One way to apply this prior knowledge is to search paths with the isotropic metric first and then apply a post-processing on the resulting graph. This strategy is straightforward but error-prone. See Fig. 2 for an example. A more efficient way to solve this problem is to introduce a smoothness constraint into the weighting function as:

$$\mathcal{W}(\ell) = \sum_{i=0}^{N-1} d(v_i)\frac{\mathcal{S}(v_i) + \mathcal{S}(v_{i+1})}{2}\mathcal{K}(v_i), \tag{2}$$

where term $\mathcal{K}(v_i)$ is a smoothness measurement of path segment defined as

$$\mathcal{K}(v_i) = \sum_{t=0}^{i} \gamma^t k_t = \sum_{t=0}^{i} \gamma^t \frac{\overrightarrow{v_{t-1}v_t} \cdot \overrightarrow{v_t v_{t+1}}}{|\overrightarrow{v_{t-1}v_t}||\overrightarrow{v_t v_{t+1}}|}, \tag{3}$$

and γ is a constant determining the influence scale of a grid point to its neighbors. Term $\overrightarrow{pq}$ denotes the vector from p to q. With this incremental smoothness term, the orientation of the previous path flow and the bending force will be involved to

determine the cost of the new extension. With the anisotropic weighting function in Eq. (2), the path detected between each pair of the seeds would be a short and smooth path riding on regions with similar properties (the foreground in our case). The total cost of the path is defined as the linking strength between the pair of seeds.

Computing the Optimal Tree. Once the linkage paths between each pair of seeds are determined, a weighted graph $\mathcal{N}(S, C, L)$ is constructed, where set S contains all the seeds. Set C consists of the optimal paths $\ell_{ij}(s_i, s_j \in S)$ with the associated edge weights in L which could be the corresponding linking strength computed above. Then a globally optimal tree with the minimum sum of edge distances can be discovered by the MST algorithm. In order to avoid the misconnection between close seeds with a sharp dark gap, a gap indicator function is incorporated in the the final vertex distance function as:

$$\mathcal{F}(\ell) = \alpha \mathcal{W}(\ell) + (1 - \alpha)\mathcal{D}(\ell), \tag{4}$$

where $\mathcal{D}$ is a gap indicator function defined as

$$\mathcal{D}(\ell) = 2/(1 + e^{I_m(\ell)/\bar{I}}). \tag{5}$$

Term I_m denotes the minimum intensity along the path and $\bar{I}$ is the mean intensity of the foreground. All the metrics are normalized for the combination.

In realistic applications, it cannot be guaranteed that there would be seeds in the critical locations of the neurite structures such as bifurcations or terminals. Actually, it is difficult to detect the branching locations in the 3D microscopy data due to the various size of structures, imaging resolution and noise. Instead of pursuing detecting those positions directly from the images, our method solves this challenge by considering the paths between the seeds. As demonstrated in Fig. 3, there are multiple paths near the bifurcation positions that are very close to each other. This indicates a converge in those locations. By merging those 'touching' path segments, more accurate bifurcation nodes are derived.

3 Experimental Results

We tested our method on 24 stacks consisting of three different data sets. The first data set consists of 16 3D stacks of projection neurons of adult Drosophila with GH146-Gal4 carrying UAS>CD2, y^+ >CD8-GFP and Hs-flp transgenes. The CD2 Flip-out cassette is removed with a mild heat shock in the third instar larvae and individual neurons are visualized with anit-GFP. The second testing set contains five confocal microscopy stacks of drosophila olfactory axonal projection neurons labelled with GFP. The last data set consists of three confocal stacks showing the arborizations of individual neurons with the main presynaptic sites (button-liked structures) in the fly lamina component. Most of the stacks have a clear view of a single neuron in a cluster. Figure. 4 demonstrates one example from each of the testing data sets. For all tests, we set $\gamma = 0.9$ and $\alpha = 0.7$.

Fig. 4. (Left) Projection Neuron of adjust drosophila. (Middle) Confocal section of drosophla olfactory axonal projection. (Right) Confocal section shows the aborisations of neurons in the lamina of drosophila.

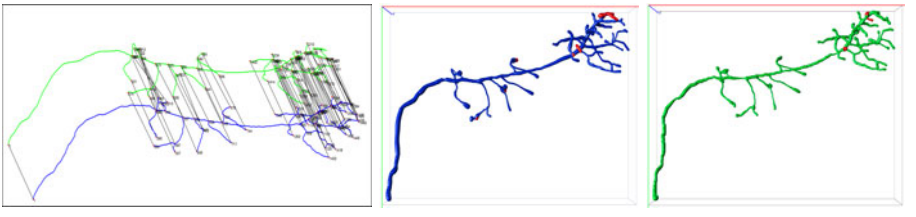


Fig. 5. An example of matching between two traces. Only bifurcation and terminal matching are demonstrated for clear view (Right). The red components in (Middle) and (Right) show the unmatched parts on the ground-truth and automatic result, respectively. In this case: MES = 0.93, $ADE(xy_d, z_d) = (5.54 \pm 1.180, 2.24 \pm 0.431)$.

To quantitatively study the performance of the tracing system, it is essential to select reliable metrics to validate the automatic traces in comparison with some gold-standard or ground-truth such as manual annotation. In our experiment, the ground-truth is generated by different people or by the same person at different times with the aid of the visualization tool V3D [10] . Instead of applying the average euclidean distance in [11], which is intuitive but vulnerable to the biocomplexity, we pursue the correspondences between key nodes in the neurite structure (such as branchings and terminals) that are of the major interest. In particular, for one node p_i, consider the set of paths in the trace originating from p_i to all other nodes. These paths provide a rich description of the entire topology of the trace relative to p_i. Assuming there are sufficient nodes in the trace, this representation is highly precise.

For computational efficiency, two features are calculated to describe each path ℓ_{ij}, including the accumulated length and the included angle between vector $\overrightarrow{p_i p_r}$ and $\overrightarrow{p_i p_j}$, where p_r refers to the root node. The statistics of those two features is a histogram h_{p_i} counting the relative locations of the other nodes regarding to p_i. Then given two different traces $\mathcal{T}_1$ and $\mathcal{T}_2$, the matching cost $\mathcal{C}(p, q)$ of two nodes $p \in \mathcal{T}_1$ and $q \in \mathcal{T}_2$ is estimated by comparing their descriptors using the χ^2 test. In all the experiments demonstrated in this paper, we choose 20 bins for included angle and 10 bins for path length respectively. Therefore, the total number of bins for each histogram is $M = 200$. This descriptor is robust since it is translation and scale invariant. In order to maintain the spatial ordering of those nodes, the

Table 1. The Evaluation on Automatic Tracing for Different Data Sets

Dataset	Avg. Size	# Stacks	MES(std.)	XY-dist(std.)	Z-dist(std.)	Time(s)
1	512x512x70	16	0.87(0.073)	3.42(1.402)	1.18(0.403)	24.6
2	512x512x73	5	0.86(0.042)	2.36(0.723)	0.76(0.181)	28.1
3	1024x1024x32	3	0.99(0.011)	1.04(0.867)	0.45(0.588)	56.2
Overall	N/A	24	0.89(0.074)	2.90(1.194)	1.00(0.380)	29.3

dynamic programming is applied to find the optimal matching π which minimizes the total matching cost: $\mathcal{Z}(\pi) = \sum_{\substack{p\in\mathcal{T}_1 \\ \pi(p)\in\mathcal{T}_2}} \mathcal{C}(p, \pi(p))$, where $\mathcal{C}(p, 0) = \eta$ is the skipping penalty which can be an instant or be determined based on the size of lost branches (directly from p). By maintaining the spatial ordering, crossing or overlapping matches are avoided. The resulting corresponding critical nodes then serve as a set of landmarks, partitioning the trace into independent segments for further analysis. Fig. 5 illustrates an example of the matching results.

With the established correspondences, we calculate two metrics to evaluate the automated tracing results. One is the Miss-Extra-Score (MES) which is defined as $\text{MES} = (S_G - S_{miss})/(S_G + S_{extra})$, where S_G is the total length of all segments in the Ground-Truth trace, S_{miss} and S_{extra} are the total lengths of missing and extra segments in the automatic trace respectively (compared with the Ground-Truth). MES provides a global view on how many parts of the neuron have been traced and how many undesired components are introduced by the automatic tracing program.

Another metric is the Average-Displacement-Error (ADE) which is defined as the average displacement of those matched components. It is a numerical measurement to describe the local accuracy of the automatic trace. ADE is valid based on the assumption that the ground-truth and automatic trace have the same root position. Because the resolution on Z-dimension is usually lower than that on XY-dimension in microscopy data, the measurement is calculated in xy- and z- dimension separately.

The results on each tested stack are plotted in Fig. 6. The majority of failure cases are found in data with blur or broken dendritic fragments, either due to insufficient dye penetration or an imaging deficiency. Table 1 summarizes the

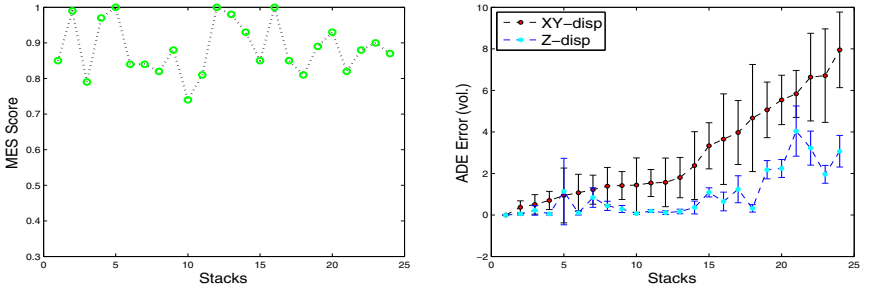


Fig. 6. The comparisons between automatic traces with the ground-truth. (Left) The MES score. (Right) The ADE error measured in XY and Z dimension respectively.

overall performances on the three testing data sets. Compared to the manual annotations, the automated 3D tracing system yielded overall MES of 0.89. The overall ADE is 2.90 and 1.00 on xy- and z- dimension respectively. The computation time varies with the stack size and the density of the neuron network. For the testing data, the average computation time is about 30s. Our system is implemented in C++ and the average computation time is obtained on a Mac Pro with 2.8GHz intel Xeon CPU and Leopard operation system.

4 Conclusion

We have developed an automated system capable of tracing neuronal structures in 3D microscopy data. Incorporating anisotropic path searching and merging, this approach is efficient for neuron tracing and bifurcation localization. In addition, a robust method is introduced for trace comparison, which can also be used for neuron recognition and structure analysis in other applications. As demonstrated in the paper, the proposed method can be applied to a variety of biomedical image tracing problems because of the robustness of global optimality and the flexibility to design a problem-specific objective function.

Acknowledgments. We thank Aljoscha Nern for generating the presynaptic site brain images and Margaret Jefferies for help of text editing the manuscript.

References

1. Cohen, A.R., Roysam, B., Turner, J.N.: Automated tracing and volume measurements of neurons from 3d confocal fluorescence microscopy data. Journal of Microscopy 173(1), 103–114 (1994)
2. Koh, I.Y.Y., et al.: An image analysis algorithm for dendritic spines. Neural Comput. 14(6), 1283–1310 (2002)
3. Shen, H., et al.: Optimal scheduling of tracing computations for real-time vascular landmark extraction from retinal fundus images. IEEE Trans. Inform. Technol. Biomed. 5, 77–91 (2001)
4. Al-kofahi, K.A., et al.: Rapid automated three-dimensional tracing of neurons from confocal image stacks. IEEE Trans. Inform. Technol. Biomed. 6, 171–187 (2002)
5. He, W., et al.: Automated three-dimensional tracing of neurons in confocal and brightfield images. Microscopy 9, 296–310 (2003)
6. Aho, A.V., Ullman, J.D., Hopcroft, J.E.: Data Structures and Algorithms. Addison-Wesley, Reading (1985)
7. Zack, G.W., Rogers, W.E., Latt, S.A.: Automatic measurement of sister chromatid exchange frequency. J. Histochem Cytochem 25(7), 741–753 (1977)
8. Felzenszwalb, P.F., Huttenlocher, D.P.: Distance transforms of sampled functions. Technical report, Cornell Computing and Information Science (2004)
9. Johnson, D.B.: Efficient algorithms for shortest paths in sparse networks. J. ACM 24(1), 1–13 (1977)
10. Peng, H., et al.: V3d enables real-time 3d visualization and quantitative analysis of large-scale biological image data sets. Nature Biotechnology 28(4), 348–353 (2010)
11. Al-kofahi, K.A., et al.: Median-based robust algorithms for tracing neurons from noisy confocal microscope images. IEEE Trans. Inform. Technol. Biomed. 7, 302–317 (2003)

An Image Retrieval Approach to Setup Difficulty Levels in Training Systems for Endomicroscopy Diagnosis

Barbara André[1,2], Tom Vercauteren[1], Anna M. Buchner[3], Muhammad Waseem Shahid[4], Michael B. Wallace[4], and Nicholas Ayache[2]

[1] Mauna Kea Technologies, Paris
[2] INRIA - Asclepios, Sophia-Antipolis
[3] Hospital of the University of Pennsylvania, Philadelphia
[4] Mayo Clinic, Jacksonville, Florida

Abstract. Learning medical image interpretation is an evolutive process that requires modular training systems, from non-expert to expert users. Our study aims at developing such a system for endomicroscopy diagnosis. It uses a difficulty predictor to try and shorten the physician learning curve. As the understanding of video diagnosis is driven by visual similarities, we propose a content-based video retrieval approach to estimate the level of interpretation difficulty. The performance of our retrieval method is compared with several state of the art methods, and its genericity is demonstrated with two different clinical databases, on the Barrett's Esophagus and on colonic polyps. From our retrieval results, we learn a difficulty predictor against a ground truth given by the percentage of false diagnoses among several physicians. Our experiments show that, although our datasets are not large enough to test for statistical significance, there is a noticeable relationship between our retrieval-based difficulty estimation and the difficulty experienced by the physicians.

1 Introduction

Objective. The understanding of pathologies through the analysis of image sequences is a subjective learning experience which may be supported by modular training systems. Particularly, the early diagnosis of epithelial cancers from *in vivo* endomicroscopy is a challenging task for many non-expert endoscopists. There is a crucial need to shorten their learning curve. Our objective is to develop a modular training system for endomicroscopy diagnosis, by adapting the difficulty level according to the expertise of the physician.

The training simulator, illustrated in Fig. 1 on the right, consists in a quiz. Given a level of difficulty, a pool of endomicroscopic videos whose average difficulty matches the current level is randomly chosen from the set of the training videos. By iterating this process with increasing levels of interpretation difficulty, the physician may be able to learn faster.

State of the art in estimating interpretation difficulty. Typical studies on query difficulty estimation consider textual queries, and not image queries.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 480–487, 2010.

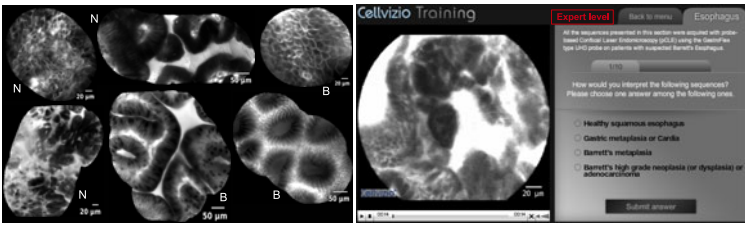


Fig. 1. Left: 6 mosaic images of *Barrett* (**B**: Benign, **N**: Neoplastic). Right: Screenshot of `www.cellvizio.net` Self-Learning Tool, with the added difficulty level information.

Besides, they usually do not predict the difficulty of the query interpretation but rather the *performance* of the query in order to estimate the quality of its retrieval results. However, given the tight analogy between text retrieval and image retrieval, the difficulty criteria used by these methods, most of which were presented in a survey by Hauff et al. [1], may also be useful for our study. In particular, Zhao et al. [2] estimated the performance of a textual query from similarity scores, but also from term frequency - inverse document frequency (TF-IDF) weights [3] extracted during the indexing time. In all these studies, the predictor validation process takes as ground truth an indicator of the performance of the retrieval system, such as the Average Precision (AP). Nevertheless, Scholer and Garcia [4] demonstrated that the correlation between the estimated difficulty and the measured retrieval performance highly depends on the chosen retrieval system. Considering human performance in rating x-ray images as a ground truth, Schwaninger et al. [5] proposed a statistical approach to estimate the image query difficulty solely from image measurements. Turpin and Scholer [6] highlighted the fact that it is not easy to establish, for simple tasks like instance recall or question answering, a significant relationship between human performance and the performance of a retrieval system that uses precision-based measures to predict the query difficulty.

For our study, we consider videos as queries. We propose to learn a query difficulty predictor using relevant attributes from a Content-Based Video Retrieval (CBVR) method. We have two types of ground truths. For video retrieval, a diagnosis ground truth is the set of histological diagnoses of the biopsies associated to all the videos of the database. For interpretation difficulty, a difficulty ground truth is given by the percentage of false video-based diagnoses among several physicians on a subset of the video database. Histological diagnosis and video-based diagnosis both consist in differentiating benign from neoplastic (i.e. pathological) lesions. In these conditions, we aim at establishing a relationship between the physicians performance and our predictor.

Materials. Probe-based confocal laser endomicroscopy (pCLE) allows the endoscopist to image the epithelial surface *in vivo*, at microscopic level with a miniprobe, and in real-time (12 frames per second) during an ongoing endoscopy.

The first pCLE database is of colonic polyps videos acquired by physicians at the Mayo Clinic in Jacksonville, Florida, USA. 68 patients underwent a

surveillance colonoscopy with pCLE for fluorescein-aided imaging of suspicious colonic polyps before their removal. For each patient, pCLE was performed of each detected polyp with one video corresponding to each particular polyp. Our resulting *Colon* database is composed of 121 videos (36 benign, 85 neoplastic) split into 499 stable video sub-sequences (231 benign, 268 neoplastic). 11 endoscopists, among whose 3 experts and 8 non-experts, individually established a pCLE diagnosis on 63 videos (18 benign, 45 neoplastic) of the database. On the non-expert diagnosis database, interobserver agreement was assessed in the study of Buchner et al. [7], with an average accuracy of 72% (sensitivity 82%, specificity 53%). On the expert diagnosis database, Gomez et al. [8] showed an interobserver agreement with an average accuracy of 75% (sensitivity 76%, specificity 72%). Thus, although pCLE is relatively new to many physicians, the learning curve pattern of pCLE in predicting neoplastic lesions was demonstrated with improved accuracies in time as observers' experience increased.

The second pCLE database is related to a different clinical application, namely the Barrett's Esophagus, and was provided by the multicentric "DONT BIOPCE" [9] study (Detection Of Neoplastic Tissue in Barrett's esophagus with In vivO Probe-based Confocal Endomicroscopy). Our resulting *Barrett* database includes 76 patients and contains 123 videos (62 benign, 61 neoplastic) split into 862 stable video sub-sequences (417 benign, 445 neoplastic). 21 endoscopists, among whose 9 experts and 12 non-experts, individually established a pCLE diagnosis on 20 videos (9 benign, 11 neoplastic) of the database.

For all these training videos, the pCLE diagnosis, either benign or neoplastic, is the same as the *gold standard* established by a pathologist after the histological review of biopsies acquired on the imaging spots.

2 Estimating the Interpretation Difficulty

For difficulty estimation, our ground truth is given by the percentage, for each query video, of false diagnoses among the physicians.As the understanding of video diagnosis by the physicians is driven by the observation of visual similarities between the query video and training videos, it makes sense to predict the query difficulty based on similarity results of video retrieval.

To learn a difficulty predictor, our idea is to exploit, as relevant attributes, the results of our video retrieval method applied to the training database. Potential relevant attributes are the class $c_q \in \{-1, +1\}$ of the video query q, the classes $c^{i\in\{1,k\}} \in \{-1, +1\}$ of its k nearest neighbors and the similarity distances $\delta^{i\in\{1,k\}}$ to them. Given the small number of videos tested by the involved physicians, too many attributes for difficulty learning may lead to over-fitting. For this reason, we decided to extract one efficient and intuitive difficulty attribute α from the retrieval results. For each query video, we considered the retrieval error between the average of the neighbors' votes and the class of the query: $\alpha_q = 1 - c_q(\sum_{i\in\{1,k\}} c^i w_{c^i})/(\sum_{i\in\{1,k\}} w_{c^i})$, where $w_{-1} = 1$ and w_{+1} is a constant weight applied to the neoplastic votes. Introducing w_{+1} allows us to take into account the possible emphasis of neoplastic votes with respect to the benign

votes. Our query difficulty predictor P is thus defined as $P(q) = \alpha_q$ for each query video q. Its relevance can be evaluated by a simple correlation measure between the estimated difficulties of all tested videos and their ground truth values. In this case, as there is no learning process, cross-validation is not necessary.

3 Video Retrieval Method

As one of the most popular method for Content-Based Image Retrieval (CBIR), the Bag-of-Visual-Words (BoW) method presented by Zhang et al. [10] aims at extracting a local image description that is both efficient to use and invariant with respect to viewpoint changes and illumination changes. Its methodology consists in first finding and describing salient local features, then in quantizing them into K clusters named visual words, and in representing the image by its signature which is the histogram of visual words. The similarity distance between two images is then defined as a distance, e.g. χ^2, between their signatures.

As the field-of-view (FOV) of single images may sometimes be insufficient to establish a diagnosis, we revisited the standard BoW method in a preliminary study [11] to retrieve videos, and not only single images. We consider each video as a set of stable sub-sequences corresponding to a relatively smooth movement of the pCLE probe along the tissue surface. We then use a video-mosaicing technique to project the temporal dimension of each sub-sequence onto one mosaic image with a larger FOV and of higher resolution. Thus, each video is considered as a set of mosaic images.

We adapt the BoW method to retrieve endomicroscopic mosaic images. In colonic polyps, a mesoscopic crypt and a microscopic goblet cell both have a rounded shape, but are different objects characterized by their different sizes. Our description should thus not be invariant with respect to scaling. Noticing that discriminative information is densely distributed in pCLE mosaic images, we decided to apply a dense detector made of overlapping disks of constant radius on a regular grid. For the description step, we used the Scale Invariant Feature Transform (SIFT) descriptor, whose combination with our dense detector keeps all the BoW related invariants except scale invariance. After mosaic image description, we define the signature of a video as the normalized sum of the visual word histograms associated to each of the constitutive mosaic images. Fig. 2 shows a CBVR example, where the retrieved videos are not only similar but also unblinded, i.e. displayed along with their contextual and diagnostic information.

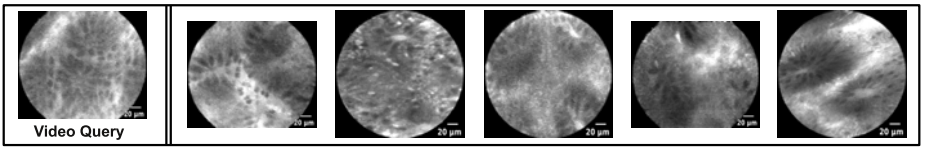


Fig. 2. Retrieval example of our CBVR method on the *Colon*. From left to right: the blinded query video and its 5 most similar videos, represented by single frames.

4 Evaluation of the Relevance of the Retrieval Method

Evaluating the relevance of retrieval results is a difficult problem. Because of the subjective appreciation of visual similarities, it is quite difficult to have a ground-truth. A simple evaluation method is to perform a classification based on the retrieval results and to estimate its accuracy. As our focus is on retrieval and not on classification, we chose one of the most straightforward classification method, the k-NN method, even though any other method could be easily plugged in. We consider two classes, benign (vote $= -1$) and neoplastic (vote $= +1$).When considering k neighbors for a query, we compute the value of the weighted sum of their votes according to their similarity distance to the query, and we compare this value with an absolute threshold θ to classify the query as benign or neoplastic. θ can be used to arbitrate between sensitivity and specificity.

In the current work, we apply for the first time our video retrieval method to two different pCLE databases, *Colon* and *Barrett*. Given their relatively small sizes, we use for each of them the whole database both for training and testing. If we only perform a leave-one-out cross-validation, the independence assumption is not respected because there are several videos acquired from the same patient. Since this may cause bias, we chose to perform a leave-one-patient-out (LOPO) cross-validation: all videos from a given patient are excluded from the training set in order to be then tested as queries of our retrieval and classification methods.

For method comparison, we will take as references the following CBIR methods, which we extended to CBVR by applying our signature summation technique: the HH-SIFT method presented by Zhang et al. [10] a sparse detector, the standard approach of Haralick features, the texture retrieval Textons method of Leung and Malik [12], and an efficient image classification method presented by Boiman et al. [13], referred as "BruteForce", that uses no clustering.

For our retrieval method, we considered disk regions of radius 60 pixels for the *Colon* database, and 20 pixels for the *Barrett* database whose discriminative patterns appear at a finer scale. We then chose 20 pixels of grid spacing to get a reasonable overlap between adjacent regions and thus be nearly invariant with respect to translation. For the number K of visual words provided by the K-Means clustering, among the values from 10 to 30000 in the literature, we chose the value $K = 100$ whose performance appeared to be sufficient for our needs.

The accuracy results of video classification on *Barrett* are presented in Fig. 3 on the left. In agreement with the presented ROC curves, the accuracy results obtained on *Colon* are even better. Our retrieval method outperforms all the compared methods with a gain of accuracy greater than 12 percentage points (pp.) on *Colon*, and greater than 9 pp. on *Barrett*. McNemar's tests show that, when the number k of neighbors is fixed, the improvement of our method with respect to all others is statistically significant: p-value < 0.011 for $k \in [1, 10]$ on *Colon* and p-value < 0.043 for $k \in [1, 2] \cup [4, 8]$ on *Barrett*. This shows the genericity of our retrieval method, which is successfully applied to two different clinical application, with: 93.4% of accuracy (sensitivity 95.3%, specificity 88.9%) at $k = 3$ neighbors on the *Colon* database, and 85.4% of accuracy (sensitivity 90.2%, specificity 80.7%) at $k = 7$ neighbors on the *Barrett* database.

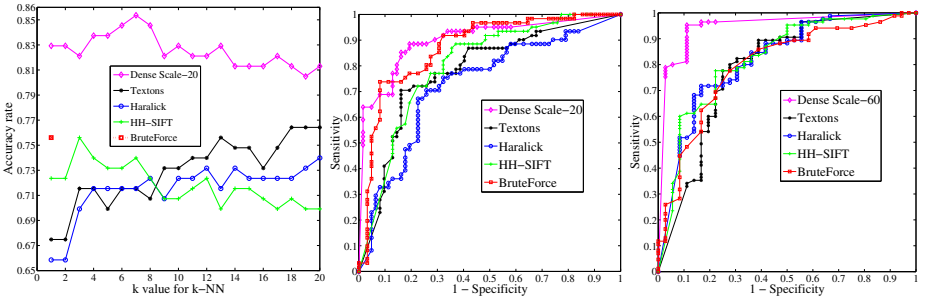


Fig. 3. Left: Method comparison for the LOPO classification of pCLE videos with $\theta = 0$ on *Barrett*. Middle and right: Corresponding ROC curves at $k = 5$ neighbors, on *Barrett* (middle) and on *Colon* (right).

5 Results of the Difficulty Estimation Method

Results on the *Barrett* database. We experimented our difficulty predictor presented in Section 2 on the *Barrett* database. The best correlation results were obtained with $k = 10$ neighbors and a neoplastic weight $w_{+1} = 0.4$. The correlation values reach 0.78 when learning from the subset of videos diagnosed by all the physicians, 0.63 when learning only from the experts and 0.80 when learning only from the non-experts. The corresponding joint histogram is presented in Fig. 4, along with the histogram of the difficulty ground truth values. We observe a noticeable relationship between ground truth and our proposed difficulty estimation, which confirms the efficiency of our retrieval-based attribute for intuitive difficulty estimation.

Perspectives for the *Colon* database. On the *Colon* database, the difficulty estimation results are not as good as on the *Barrett* database. With $k = 10$ neighbors and a neoplastic weight $w_{+1} = 6$, the correlation values reach 0.45

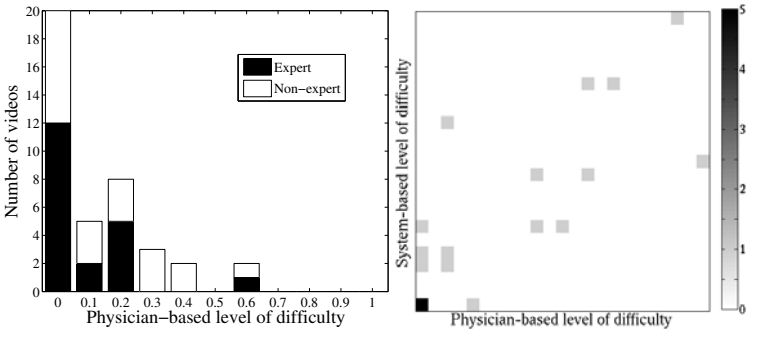


Fig. 4. Left: Difficulty ground truth histogram on *Barrett*. Right: Joint histogram, x-axis is the difficulty of all the physicians and y-axis is our estimated difficulty. 21 physicians, 9 expert and 12 non expert, individually diagnosed 20 videos.

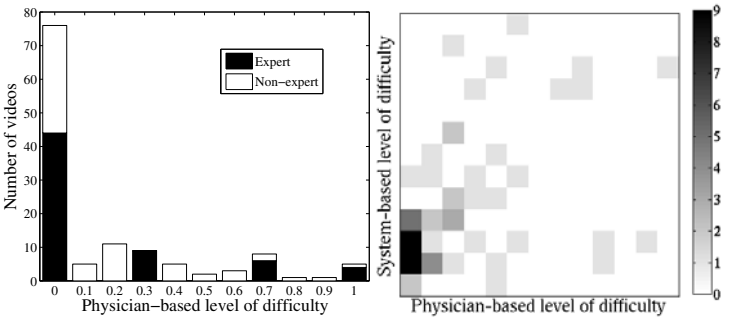


Fig. 5. Left: Difficulty ground truth histogram on *Colon*. Right: Joint histogram, x-axis is the difficulty of all the physicians and y-axis is our estimated difficulty. 11 physicians, 3 expert and 8 non expert, individually diagnosed 63 videos.

when learning from the subset of videos diagnosed by all the physicians, 0.30 when learning from experts and 0.45 when learning from non-experts.

In order to improve these results, we propose to investigate a machine learning-based approach, which will need more relevant attributes. As the video dataset for which we have the difficulty ground truth is relatively small, we decided to add one discriminative power attribute β and to learn the difficulty predictor from the two attributes α and β by using a robust linear regression model. Our discriminative power attribute β reflects the deviation of the "signed" discriminative power of the query signature, with respect to the benign and the neoplastic classes: $\beta = std(\sum_{i\in\{1,K\}} f_i d(i))$. In this formula, the visual word i has a frequency f_i in the video query and its "signed" discriminative power $d(i)$ is given by the adapted Fisher criterion: $d(i) = (\mu_{-1} - \mu_{+1})|\mu_{-1} - \mu_{+1}|/(0.5\ (\sigma^2_{-1} + \sigma^2_{+1}))$, where μ_c and σ_c are respectively the mean and the variance of the frequency distribution of the visual word i in the videos belonging to class c.

The correlation values obtained by the robust linear regression model with cross-validation reach 0.48 when learning from the subset of videos diagnosed by all the physicians, 0.33 when learning only from the experts and 0.47 when learning only from the non-experts. Even if these correlation results are less convincing than those obtained on the *Barrett* database, the correlation tendency can be qualitatively appreciated. The corresponding joint histogram is presented in Fig. 5. To automate the optimal attributes selection and to explore more potentially relevant attributes for difficulty estimation, further experiments based on model selection need to be investigated, for example using the Akaike information criterion. Besides, selection criteria commonly used in active learning [14] may help to provide a better difficulty estimation.

6 Conclusion

To our knowledge this study proposes the first approach to estimate interpretation difficulty for endomicroscopy training, based on an original method of Content-Based Video Retrieval. Our experiments have demonstrated that there is a noticeable relationship between our retrieval-based difficulty estimation and the difficulty experienced by the physicians. Moreover, we showed the promising

genericity of our difficulty estimation method by applying it on two different clinical databases, one on the Barrett's Esophagus and the other on colonic polyps. Our method could also be potentially applied to other imaging applications.

On one hand we have the diagnosis ground truth for all the videos belonging to our two large databases, on the other hand we have the difficulty ground truth on a small subset of each database. The method proposed in this work can then be used to estimate the interpretation difficulty on the remaining videos. The complete databases could thus be used in a training simulator that features difficulty level selection. This should make endomicroscopy training more relevant. Finally, a clinical validation would be required to see whether such a structured training simulator could help shorten the physician learning curve.

References

1. Hauff, C., Hiemstra, D., de Jong, F.: A survey of pre-retrieval query performance predictors. In: Proc. CIKM 2008, pp. 1419–1420 (2008)
2. Zhao, Y., Scholer, F., Tsegay, Y.: Effective pre-retrieval query performance prediction using similarity and variability evidence. In: Macdonald, C., Ounis, I., Plachouras, V., Ruthven, I., White, R.W. (eds.) ECIR 2008. LNCS, vol. 4956, pp. 52–64. Springer, Heidelberg (2008)
3. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. In: Information Processing and Management, pp. 513–523 (1988)
4. Scholer, F., Garcia, S.: A case for improved evaluation of query difficulty prediction. In: Proc. SIGIR 2009, pp. 640–641 (2009)
5. Schwaninger, A., Michel, S., Bolfing, A.: A statistical approach for image difficulty estimation in X-ray screening using image measurements. In: Proc. APGV 2007, pp. 123–130 (2007)
6. Turpin, A., Scholer, F.: User performance versus precision measures for simple search tasks. In: Proc. SIGIR 2006, pp. 11–18 (2006)
7. Buchner, A.M., Gomez, V., Gill, K.R., Ghabril, M., Scimeca, D., Shahid, M.W., Achem, S.R., Picco, M.F., Riegert-Johnson, D., Raimondo, M., Wolfsen, H.C., Woodward, T.A., Hasan, M.K., Wallace, M.B.: The learning curve for in vivo probe based Confocal Laser Endomicroscopy (pCLE) for prediction of colorectal neoplasia. Gastrointestinal Endoscopy 69(5), AB364–AB365 (2009)
8. Gomez, V., Buchner, A.M., Dekker, E., van den Broek, F.J., Meining, A., Shahid, M.W., Ghabril, M., Fockens, P., Wallace, M.B.: Interobserver agreement and accuracy among international experts of probe-based confocal laser microscopy (pCLE) in predicting colorectal neoplasia. Endoscopy (in press, 2010)
9. DONT BIOPCE, http://clinicaltrials.gov/ct2/show/NCT00795184
10. Zhang, J., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: a comprehensive study. Int. J. Comput. Vis. 73, 213–238 (2007)
11. André, B., Vercauteren, T., Wallace, M.B., Buchner, A.M., Ayache, N.: Endomicroscopic video retrieval using mosaicing and visual words. In: Proc. ISBI 2010 (in press, 2010)
12. Leung, T., Malik, J.: Representing and recognizing the visual appearance of materials using three-dimensional textons. Int. J. Comput. Vis. 43, 29–44 (2001)
13. Boiman, O., Shechtman, E., Irani, M. In: defense of nearest-neighbor based image classification. In: Proc. CVPR 2008, 1–8 (2008)
14. Hoi, S.C.H., Jin, R., Zhu, J., Lyu, M.R.: Semi-supervised SVM batch mode active learning for image retrieval. In: Proc. CVPR 2008, pp. 24–26 (2008)

3D Localization of Pronuclei of Human Zygotes Using Textures from Multiple Focal Planes

A. Giusti[1,⋆], G. Corani[1], L. Gambardella[1], C. Magli[2], and L. Gianaroli[3]

[1] Dalle Molle Institute for Artificial Intelligence (IDSIA), Lugano, Switzerland
[2] International Institute for Reproductive Medicine (IIRM), Lugano, Switzerland
[3] INFERGEN, Lugano, Switzerland

Abstract. We propose a technique for recovering the position and depth of pronuclei of human zygotes, from Z-stacks acquired using Hoffman Modulation Contrast microscopy. We use Local Binary Pattern features for describing local texture, and integrate information from multiple neighboring areas of the stack, including those where the object to be detected would appear defocused; interestingly, such defocused areas provide very discriminative information for detection. Experimental results confirm the effectiveness of our approach, which allows one to derive new 3D measurements for improved scoring of zygotes during In Vitro Fertilization.

1 Introduction

During In Vitro Fertilization (IVF), *fertilization* is the union of a sperm with an oocyte; the resulting *zygote* contains genetic material (DNA) from both the father and the mother. Usually, several zygotes per patient are successfully fertilized, but only a limited number can be cultured (no more than three in many countries). It is therefore necessary to select the zygotes to be cultured. Currently, embryologists perform zygote selection on the basis of a morphological evaluation of the zygotes [4], which are observed through an optical microscope equipped with Hoffmann Modulation Contrast (HMC) optics.

Several scoring systems have been designed in order to recognize zygotes with high potential of implantation; yet, none seems to be strong enough to reliably predict whether an IVF cycle results in a pregnancy [11]. Within such scoring systems, the relative position of the *pronuclei* (see Fig. 1) is a factor of major importance, but so far, only 2D information has been available for measurement. Yet, the zygote has a spherical shape and thus represents a thick sample with respect to the microscope depth of field: it is therefore possible to acquire several images at different, equally-spaced focus planes. This is a common practice in automated microscopy, and the resulting set of images is named focus stack (or Z-stack): in each of the images of the stack, structures lying at the corresponding depth appear in-focus, whereas other structures appear blurred.

⋆ Contact: alessandrog@idsia.ch – Additional material at http://www.idsia.ch/~giusti/IN3

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 488–495, 2010.

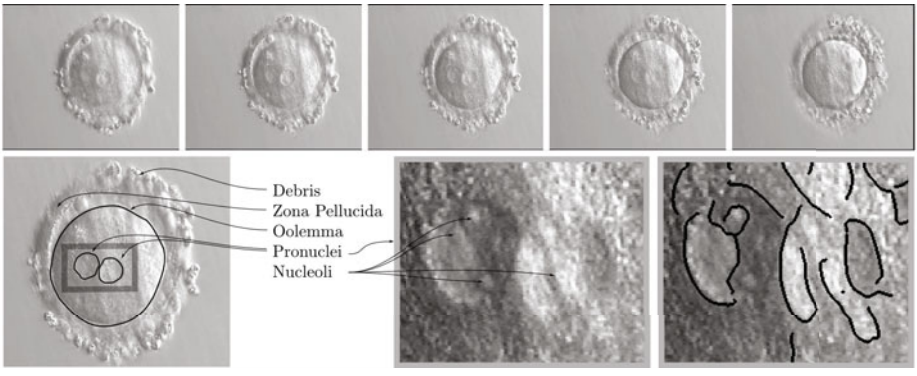


Fig. 1. A Zygote. *Top row*: 5 of the 24 images in the source focus stack. Note that different structures are visible in different images, depending on their depth. *Bottom*: parts of the zygote, and detail of pronuclei appearance, when they are both in sharp focus: this is possible because they lie at the same depth in this specific case. Right: edge detection, even with manually-tuned parameters, fails due to weak borders.

In this paper we present a technique to localize pronuclei in a focus stack, determining both their 2D position and depth; this allows us to perform a partial 3D reconstruction of the zygote morphology, from which a larger and more consistent array of measurements can be performed.

Detecting pronuclei in HMC image stacks is challenging because of their wide appearance variability (see Figure 2) and large amounts of clutter. We exploit Local Binary Patterns (LBP) [7] as a texture feature for pronuclei detection, thus leveraging on the peculiar (albeit complex and unpredictable) appearance of the pronucleus' interior. Additionally, our detector integrates information from several neighboring focal planes: our experimental findings show that even texture data from focal planes where the pronucleus appears defocused provide informative features.

Images from human embryology have only recently been subject to attention from the image processing community, because of their complex appearance, and the limited need of processing large amounts of data, as samples were observed only once per day. The availability of new systems integrating microscopes in incubators [10] will produce much more frequent observations, and increase the need of automated processing: moreover, significant advantages of performing objective, quantitative measurements have been shown by Beuchat et al. in [2], where various 2D zygote features, including pronuclei positions and sizes, are measured from a single image (where all such features are assumed to be in focus) with user assistance.

In [6,5] we presented algorithms for processing image stacks, with the goal of segmenting the outer shape of the zygote, or the blastomeres in an embryo. The present work fits in the same approach of also determining the object's depth from the image stack. Unfortunately, pronuclei lie at unpredictable positions and depth, and do not provide sufficiently visible outer edges, nor apparent

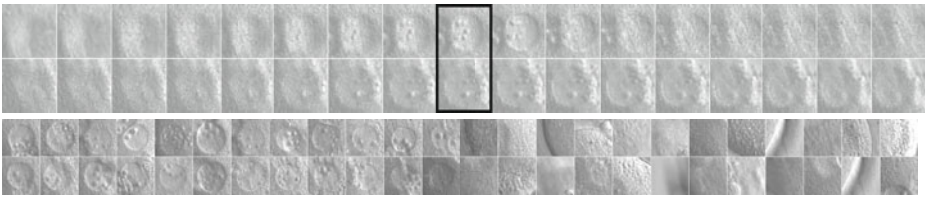


Fig. 2. *Top*: Defocused appearance of pronuclei (left and right), vs best focused plane (center). *Bottom*: positive (left) and negative (right) examples.

area features; this prevents the direct application of segmentation techniques such as active contours [3] or level sets [12], unless the pronuclei 3D location is first determined, which is the problem we are dealing with in this paper. Then, having determined a reasonable initialization, a simple snake-based approach can provide a precise segmentation and size measurement, exploiting weak large-scale gradients through strong shape priors.

Several recent works deal with the detection of cell nuclei or other subcellular structures; the problem is radically different in fluorescent-microscopy approaches or when nuclei are stained, both of which are not viable options for human zygote observation; ray features [13] have recently proven useful in capturing the irregular shapes of neuron nuclei. In our scenario their application would not be justified, as pronuclei lack clear edges but have regular shapes.

2 Texture Features at Multiple Focal Planes

HMC is a complex optical microscopy technique, which visualizes differences in optical density as variations in light intensity: it is routinely used in IVF labs for observing zygotes, as it provides a large amount of contrast for transparent specimens and eases human observation as the objects appear three-dimensional and side-lit, as if a virtual light source was illuminating them from a side (apparent lighting direction). In particular, images of pronuclei share the following appearance traits.

1. Pronuclei do not generally exhibit clear edges at any focal plane, which limits the utility of approaches based on edge detection – such as hough transform, which would perform well due to the predictable circular overall shape; they also lack a predictable large-scale appearance, which could be captured by haar-like features.
2. The interior of pronuclei has a peculiar, subtle texture when in focus; moreover, nucleoli are often visible, but have a very variable size, distribution and appearance in the pronucleus.
3. Different zygote images are subject to large photometric variations, even when imaging parameters are unchanged.
4. Zygote images are often cluttered by large amounts of high-contrast debris, as well as vacuoles and other structures mimicking the large-scale pronuclei appearance.

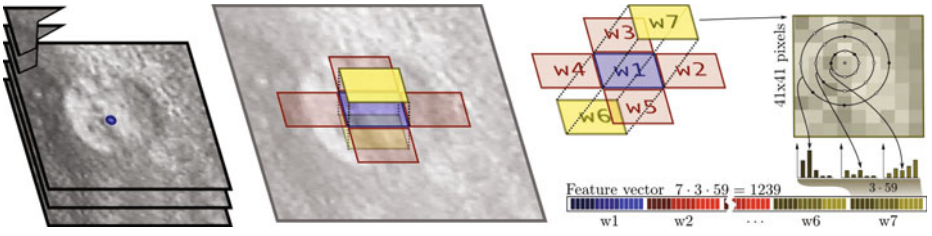


Fig. 3. Composition of the feature vector, which is obtained from 8-neighbor LBP features computed in 7 different windows at 3 scales

We use Local Binary Patterns (LBP) features for dealing with this challenging scenario [7]. Such descriptor is amongst the most powerful options for texture analysis and has the advantage of simple implementation and fast performance, and has been successfully used in several biomedical applications [14,8]. LBP features allow us to represent the characteristic interior texture of the zygote, and, as the descriptor is histogram-based, it captures the presence of nucleoli regardless of their actual position in the pronucleus (point 2); moreover, the invariance of the descriptor to grayscale transformations make it robust to the marked unpredictability in photometric characteristics between images described in point 3 above.

In order to capture larger-scale, structural characteristics of pronuclei, we concatenate LBP descriptors computed on multiple neighboring windows; a similar approach has been used for face description [1]. For a given point $p = (x, y, z)$, we compute a feature vector by considering data from 7 different 41x41 windows of the image stack. Window w1 is centered on p, whereas windows w2 to w5 are positioned around it (see Figure 3). We also consider windows w6 and w7, centered on the same (x, y) position as p, but containing data from focal planes below and above p, respectively. In each window, we compute LBP descriptors using 8 neighbors and uniform mapping, which originates a 59-bin histogram; features are computed for 3 different scales, at radii 1, 2 and 3. Feature vectors for each of the 3 scales at each of the 7 windows are concatenated, thus creating a global feature vector for point p composed by $7 \cdot 3 \cdot 59 = 1239$ variables, which summarizes data from multiple focal planes. We will then use statistical classification to predict, on the basis of these 1239 variables, whether point p belongs or not to a pronucleus.

Extensions of LBP features to a third dimension have been successfully proposed for several applications [17], mostly dealing with dynamic content. One common approach is to consider the texture characterizing $x-t$ and $y-t$ space-time slices (LBP-TOP). In our evaluations, texture in $x-z$ or $y-z$ slices proved uninformative in dealing with focus stacks, regardless of the scale factor along the z axis; this is probably due to marked anisotropy in such slices because of the effects of defocus: therefore, we only include descriptors for $x-y$ planes in our final feature vector.

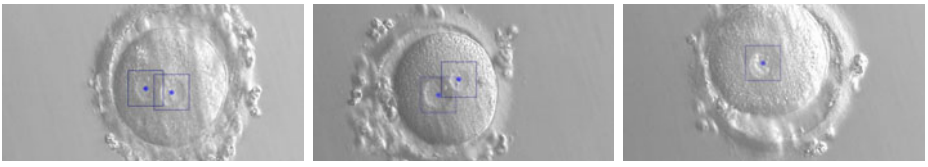


Fig. 4. Detection results

3 Experiments

We experimented our technique on 193 focus-stacks representing zygotes, acquired during routine IVF observations using an IX51 microscope equipped with 20x HMC optics and a 720x576 camera. Each stack is composed by 24 frames, acquired at focal planes with a z spacing of roughly $5\mu m$. Zygotes and pronuclei have an apparent diameter of approximately 300 and 50 pixels, respectively. The approximate position and depth of pronuclei in each image stack is labeled by an operator, by means of an ad-hoc focus-stack viewer. In total, our dataset contains 347 labeled pronuclei (1.8 per image stack on average).

In order to detect pronuclei in a new image, feature vectors are computed for each point on a 5x5 grid, on all the images in the stack; the process is not computationally demanding, because LBP codes can be precomputed on the whole image. Processing can be significantly sped up by limiting the evaluation to points inside the zygote oolemma, which can be robustly found using the technique in [6], thus limiting to 50 seconds the time for processing a full stack in our MATLAB implementation. Each point is classified for the presence of a pronucleus by means of the classifier described in the following section. Positive responses, which constitute a small fraction of the tested points, are used to initialize snake-based detectors; this automatically handles multiple positive responses from a single pronucleus, as multiple snakes initialized near a pronucleus converge and finally overlap.

From the previously described labeled data, we built a data set of 347 positive examples (i.e., points belonging to the pronucleus) and 11927 negative examples, which are centered on random points farther than $15\mu m$ from any pronucleus center.

Classification. First, we split the data into two halves, each containing some 6000 instances; we used the *selection set* to perform feature selection and to select the classification algorithm; we used the *test set* to eventually measure the performance of the chosen classifier[1]. We performed all classification experiments using WEKA [16].

We started by computing the information gain (IG) of each feature; the IG is a standard index of predictivity for features: better features have a higher IG. As shown in Figure 5 (central plot), the distribution of IG is highly skewed,

[1] Keeping the test set separate from the data used for selecting features and classification algorithm is necessary to avoid optimistic measures of accuracy [9].

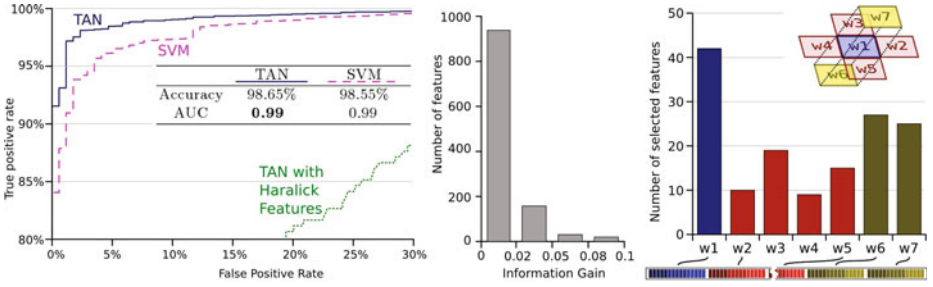


Fig. 5. *Left*: ROC curve, accuracy and area under curve (AUC) measured on the test set. *Center*: distribution of the IG across the 1239 features. *Right*: number of selected features in each window, after IG analysis.

namely few features have high IG, while most features have a very low one. We retained the features having $IG > IG_{max}/3$, where IG_{max} is the maximum IG among all features. In this way, we selected 152 features; the average IG of the selected features is about five times larger than the average IG of the non-selected features. Most selected features are from regions $w1$, $w6$ and $w7$, as shown in Figure 5 (right plot). This shows that focal planes where the pronucleus appears defocused ($w6$ and $w7$) provide useful texture information.

The IG analysis, although useful for a preliminary screening, analyzes only one feature at a time and cannot capture interaction between features. To identify and remove correlated features, we then used the the correlation-based feature selector (also available within WEKA), further reducing the features to 45.

By experiments on the selection set we select SVM with linear kernel and TAN (Bayesian network with tree-augmented structure) as the best performing classifiers; they are slightly but significantly better than ADABoost with decision stumps [15]. Eventually, we trained the classifiers on the selection set and evaluated them on the test set, measuring for both accuracy of 99% and AUC (area under the ROC curve) higher than 0.99.

Further Classification Experiments. An usually desirable property of texture features is rotation invariance, which can be obtained in LBP features by using a different mapping between LBP codes and histogram bins [7]. However, rotation-invariant features are much less powerful than those obtained with uniform mapping; the IG of the best 10 rotation-invariant features is on average 2.5 times smaller than that of the best 10 features with uniform mapping. Moreover, classification accuracy drops to 96% (both for SVM and TAN) when rotation invariant features are used; such an accuracy corresponds to that of a trivial classifier, which always returns the label "negative".

We compared the LBP-based approach described in this paper to Haralick texture features computed in a single 40x40 window around the candidate point; features are computed on normalized intensities using 8 equally-spaced intensity bins, and considering two orthogonal displacements; such naive approach yielded significantly worse results in terms of both IG (about four times smaller for the

10 best Haralick features compared to best 10 features with our LBP-based descriptor) and classification accuracy (which drops to 96%); figure 5 (left) reports a dotted line for the resulting ROC curve.

We investigated to what extent the high accuracy of the classifiers shown in Figure 5 was due to the unbalance of the data set; in fact, a classifier always returning the class "negative" would have achieved an accuracy of 96%. We therefore built a data set of some 700 examples, evenly distributed between positive and negative. Both TAN and SVM achieved 95% accuracy, thus confirming the high predictivity of LBF from multiple focal planes (recall that in this case, simply returning "negative" would only achieve 50% accuracy).

Localization and Segmentation. In order to precisely localize pronuclei and determine their size, from each positively-classified point we initialize a single circular active contour, with three DOF for position and one additional DOF for size. Such naive low-dimensional model is evolved using gradient descent in order to minimize an energy measure, determined from large-scale image gradients near its border, which are expected to have directions consistent with the apparent lighting due to HMC, as described in [6]. Candidate pronuclei are determined by solutions not overlapping in 3D space more than 30% of their volume, and amount to an average of 3.4 per image.

We consider a candidate being correctly segmented if three conditions are met: (a) its center is within 3 μm (10 pixels) of the manually-labeled center; (b) its depth less than 2 stack slices away from the labeled depth – which is the depth where the pronucleus edges appear sharpest; (c) its radius not differing more than 2 μm from the actual one. We visually determined that circular models meeting these constraints manage to exactly fit the pronucleus shape in all instances, and routinely exceed the precision of manual labeling of the center of pronuclei. This prevents us from quantifying localization error, as exact ground truth is not available.

In our dataset, 93.3% of pronuclei are correctly segmented and appear in the final list of candidates; in our present implementation, candidates are displayed to users during routine evaluation of the zygote image stacks, and are manually confirmed with a single click on each – after which the related position, depth and size measurements are saved in the database and used as additional features for statistical evaluations. Naively selecting the two lowest-energy candidates provides correct results for both pronuclei in 84.0% of our image stacks, whereas in 14.3% of the instances only one of the two pronuclei is correctly detected.

4 Conclusions

We presented a technique for detecting pronuclei in focus stacks, and determining their depth, by exploiting texture information from several focal planes simultaneously. Experimental results show promising results, which enable users to perform zygote scoring with an insight of the zygote's 3D morphology. As future works, we plan to deal with timelapse data, which poses a new set of challenges for automated processing.

Acknowledgements. Research partially supported by CTI (Confederation's innovation promotion agency) grant n. 9707.1 PFSL-LS and NSF 200021-118071/1.

References

1. Ahonen, T., et al.: Face description with local binary patterns: application to face recognition. Trans. on PAMI (2006)
2. Beuchat, A., Thévenaz, P., Unser, M., Ebner, T., Senn, A., Urner, F., Germond, M., Sorzano, C.O.S.: Quantitative morphometrical characterization of human pronuclear zygotes. Human Reproduction 23, 1983–1992 (2008)
3. Blake, A., Isard, M.: Active Contours. Springer, Heidelberg (1998)
4. Gianaroli, L., Magli, M.C., Ferraretti, A.P., Lappi, M., Borghi, E., Ermini, B.: Oocyte euploidy, pronuclear zygote morphology and embryo chromosomal complement. Human Reproduction 22(1), 241–249 (2007)
5. Giusti, A., Corani, G., Gambardella, L., Magli, C., Gianaroli, L.: Blastomere segmentation and 3d morphology measurements of early embryos from hoffman modulation contrast image stacks. In: Proc. of ISBI (2010)
6. Giusti, A., Corani, G., Gambardella, L., Magli, C., Gianaroli, L.: Simultaneous focusing and contouring of human zygotes for in vitro fertilization. In: Proc. of BIOSIGNALS (2010)
7. Ojala, T., et al.: Gray scale and rotation invariant texture classification with local binary patterns. In: Vernon, D. (ed.) ECCV 2000. LNCS, vol. 1842, pp. 404–420. Springer, Heidelberg (2000)
8. Qureshi, H., et al.: Adaptive discriminant wavelet packet transform and local binary patterns for meningioma subtype classification. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 196–204. Springer, Heidelberg (2008)
9. Reunanen, J.: Overfitting in making comparisons between variable selection methods. The Journal of Machine Learning Research 3, 1371–1382 (2003)
10. Scott, L., et al.: Innovative techniques in human embryo viability assessment. Reproductive Biomedicine online (2008)
11. Scott, L., Finn, A., O'Leary, T., McLellan, S., Hill, J.: Morphologic parameters of early cleavage-stage embryos that correlate with fetal development and delivery: prospective and applied data for increased pregnancy rates. Human Reproduction 22(1), 230–240 (2007)
12. Sethian, J.A.: Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge (1999)
13. Smith, K., et al: Fast ray features for learning irregular shapes. In: Proc. of CVPR (2009)
14. Sørensen, L., et al.: Texture classification in lung ct using local binary patterns. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 934–941. Springer, Heidelberg (2008)
15. Viola, Jones: Rapid object detection using boosted cascade of simple features. In: Proc. of CVPR (2001)
16. Witten, I.H., Frank, E.: Data Mining: Practical Machine Learning Tools and Techniques, 2nd edn. Morgan Kaufmann, San Francisco (2005)
17. Zhao, G., Pietikäinen, M.: Dynamic texture recognition using local binary patterns with an application to facial expressions. Trans. on PAMI (2007)

Motion Compensated SLAM for Image Guided Surgery

Peter Mountney and Guang-Zhong Yang

Department of Computing and Institute of Biomedical Engineering
Imperial College, London SW7 2BZ, UK
{peter.mountney05,g.z.yang}@imperial.ac.uk

Abstract. The effectiveness and clinical benefits of image guided surgery are well established for procedures where there is manageable tissue motion. In minimally invasive cardiac, gastrointestinal, or abdominal surgery, large scale tissue deformation prohibits accurate registration and fusion of pre- and intra-operative data. Vision based techniques such as structure from motion and simultaneous localization and mapping are capable of recovering 3D structure and laparoscope motion. Current research in the area generally assumes the environment is static, which is difficult to satisfy in most surgical procedures. In this paper, a novel framework for simultaneous online estimation of laparoscopic camera motion and tissue deformation in a dynamic environment is proposed. The method only relies on images captured by the laparoscope to sequentially and incrementally generate a dynamic 3D map of tissue motion that can be co-registered with pre-operative data. The theoretical contribution of this paper is validated with both simulated and *ex vivo* data. The practical application of the technique is further demonstrated on *in vivo* procedures.

Keywords: Image Guided Surgery, Minimally Invasive Surgery, Tracking, Simultaneous Localization And Mapping (SLAM), Augmented Reality.

1 Introduction

For Minimally Invasive Surgery (MIS), the use of pre- and intra-operative image guidance has well established benefits. However, its application to procedures with large tissue deformation, such as those encountered in cardiovascular, gastrointestinal and abdominal surgery, is still limited. In order to establish a common *in vivo* material frame of reference that follows tissue deformation, *in situ* 3D reconstruction is necessary. In this regard, the use of fiducial markers and optical tracking, as well as intra-operative imaging such as ultrasound, MR and x-ray fluoroscope have been explored extensively. However, the use of vision techniques based on images from laparoscopes/endoscopes during MIS has clear advantages. It does not require the introduction of additional equipment to what is already a very complex surgical setup. Furthermore, it defines a single co-ordinate system for intra-operative 3D reconstruction, imaging device localization and visualization, therefore removing the need for registration between multiple data streams to a global coordinate system.

The vision based techniques used for MIS currently include Simultaneous Localization And Mapping (SLAM) [1, 2] and Structure from Motion [3, 4]. They have been applied to a variety of anatomical settings such as the abdomen [1, 2],

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 496–504, 2010.

colon [3], bladder [4] and sinus [5], with the assumption that the structure is relatively static. Structure from Motion has been formulated for used in non-rigid environments however, it requires offline batch processing, thus making it difficult for real-time applications. In [6], for example, it is used to estimate a static cardiac surface at a pre selected point in the cardiac cycle. The 3D structure of the deforming cardiac surface is estimated online in [7] by tracking regions of interest on the organ. It is important to note that in approaches such as this, it is assumed that the laparoscopic camera is fixed, which is not realistic for *in vivo* applications.

The purpose of this paper is to present a novel online approach for simultaneous estimation of camera motion and deforming tissue structure. The system presented extends the current SLAM framework, not only to cope with camera motion, but also to learn a high level model for compensating periodic organ motion. The learnt motion model is explicitly incorporated into the statistical SLAM framework, enabling dynamic tissue motion to be estimated even when it is outside the camera's current field-of-view. The basic steps of the proposed algorithm is schematically illustrated in Fig 1, those specific steps for dealing with dynamic tissue motion are highlighted in yellow. We term this Motion Compensated SLAM (MC-SLAM), which to our knowledge, is the first work for simultaneous online estimation of camera motion and dynamic structure. To assess the accuracy of the proposed framework, the proposed method is validated with synthetic and *ex vivo* data and its *in vivo* application is also demonstrated.

2 Methods

2.1 Motion Modeling

To illustrate the practical use of the MC-SLAM framework, we will use MIS liver surgery as an example. It has been shown that the motion of the liver is correlated to the periodic motion of the diaphragm and therefore respiration [8]. In this work we exploit this correlation to create a high level model of respiration that can be used to predict the dynamic 3D position of tissue in the abdomen. The respiration model is estimated by measuring the 3D motion of points on the liver, as shown in Fig 1b), with a stereo laparoscope. The 3D position of points on the liver are estimated by matching regions of interest in the left and right stereo images and triangulating using the camera's intrinsic and extrinsic parameters. The temporal motion of the 3D points is estimated by tracking the regions of interests along time using an approach outlined in [9]. This approach learns what information is unique about a region and how best to distinguish it from its surroundings, making it well suited to MIS data.

In this work, the liver is assumed to move and deform freely in 3D. Fig. 2a) illustrates the 3D coordinates of a region on the liver surface. The data was collected from a static laparoscope during an in vivo porcine experiment. The signal is periodic in all three axes. The point is periodically moving along a path or principal axis in 3D space which corresponds to the superior-inferior axis [10] and can be approximated as locally linear as shown in Fig 1b). Each point on the liver has an individual principal axis of motion. However, its position on that axis is correlated to the respiration cycle.

By determining the principal axis of motion and observing the temporal motion characteristics, a model of respiration can be inferred.

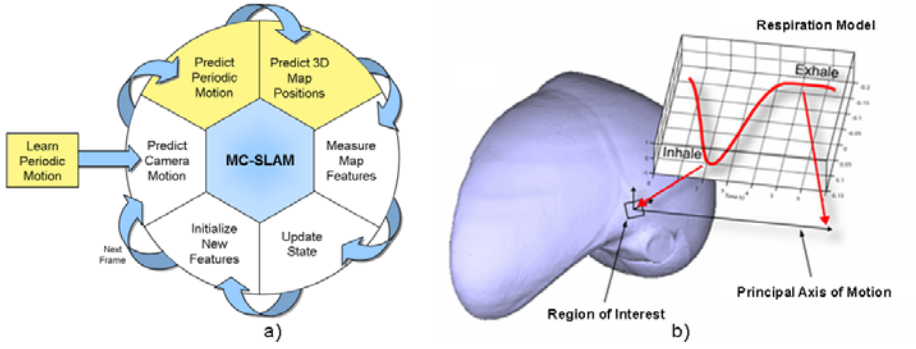


Fig. 1. (a) A schematic illustration of the main steps of the proposed MC-SLAM framework. Additional steps for dealing with dynamic tissue motion are highlighted in yellow. **(b)** An example illustration of respiratory modeling from organ motion, which involves: 1) the motion of a region or feature point (of a liver) is tracked temporally in 3D, 2) the principal axis of motion (a vector representing the dominant direction of organ motion) is estimated, 3) the periodic motion along this axis is examined and a respiration model is fitted to the data.

In order to relate the 3D coordinate space to the principal axis of motion, Principal Component Analysis (PCA) is used. The result of PCA for the data in Fig. 2a) is shown in Fig. 2b). The first component of PCA is shown in blue, which clearly represents respiration induced tissue motion. The second component contains a small variance caused by hysteresis. A typical respiratory cycle is asymmetrically periodic [10] with a longer dwell time at exhalation as shown in Fig. 2c). This can be represented as

$$z(t) = z_0 - b\cos^{2n}\left(\frac{\pi t}{\tau} - \phi\right) \tag{1}$$

where z_0 is the position of the liver at the exhale, b is the amplitude, τ is the respiration frequency, ϕ is the phase and n describes the shape or gradient of the model. Eq. 1 is used to model the data in the first component of PCA (shown in Fig. 2b). The parameters of Eq. 1 are estimated using Levenberg-Marquardt minimization algorithm where the problem is posed as a least squares curve fitting.

The respiration cycle can be estimated using any point on the liver, assuming it can be tracked throughout the respiratory cycle. The transformation from the global coordinate system to the respiration coordinate system is unique to each point. This means that points on the surface of the liver can move and deform in independent directions but share the same respiration model. Given a model of respiration, it is therefore possible to estimate the dynamic tissue motion using the inverse PCA transformation matrix and a given point in the respiration cycle.

In MIS, respiration is normally regulated by a ventilator. The respiration cycle can therefore be considered periodic with small fluctuations in the frequency and amplitude caused by the ventilator. In the following section, we show how the periodic respiration and associated ventilator noise can be modeled in an Extended Kalman Filter (EKF) to prevent error propagation and synchronization issues.

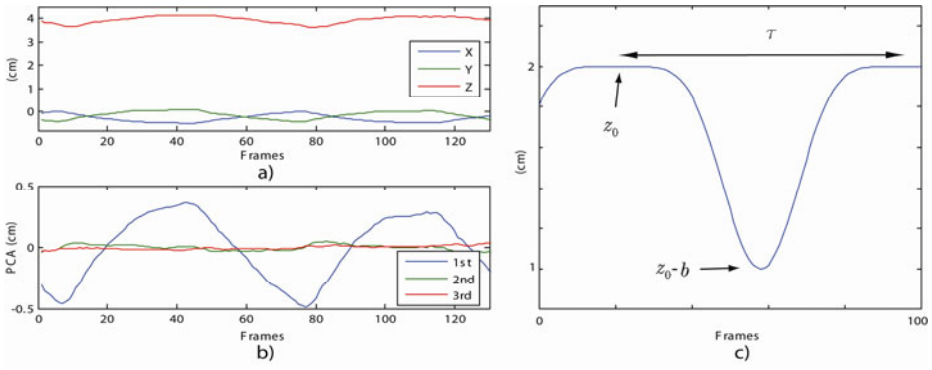


Fig. 2. (a) The 3D global coordinates of a region on the surface of the liver illustrating periodic organ motion. **(b)** Result of PCA as applied to (a), illustrating the respiration cycle extracted from organ motion in (a). The first 3 PCA components are shown and the first component corresponds to the principal axis of motion. **(c)** is a graphical representation of the asymmetric respiration model described by Eq. 1.

2.2 Motion Compensated SLAM (MC-SLAM)

In SLAM [1, 2], it is generally assumed that the map is rigid. In MC-SLAM, a periodic motion model is introduced to compensate for the dynamic tissue deformation, thus enabling dynamic mapping and accurate localization of the camera. Conceptually, this introduces three novel steps into the SLAM framework shown in Fig. 1a). The first is to learn an initial estimate of the periodic respiration model using the method described in the previous section. The second and third steps are the prediction of the respiration motion model and prediction of the tissue motion within the map. In conjunction with these steps, we have introduced a new state vector, prediction model and measurement model.

MC-SLAM's **probabilistic framework** is an Extended Kalman Filter (EKF). The state vector $\hat{x}$ is composed of three elements representing the camera $\hat{x}_v$, the periodic respiration model $\hat{m}$ and the map $(\hat{y}_1 \cdots \hat{y}_i)$. The covariance matrix P is square where $P_{\hat{x}_v\hat{m}}$ is the covariance between state elements $\hat{x}_v$ and $\hat{m}$.

$$\hat{x} = \begin{pmatrix} \hat{x}_v \\ \hat{m} \\ \hat{y}_1 \\ \hat{y}_2 \\ \vdots \end{pmatrix} \quad P = \begin{pmatrix} P_{\hat{x}_v\hat{x}_v} & P_{\hat{x}_v\hat{m}} & P_{\hat{x}_v\hat{y}_1} & P_{\hat{x}_v\hat{y}_2} & \cdots \\ P_{\hat{m}\hat{x}_v} & P_{\hat{m}\hat{m}} & P_{\hat{m}\hat{y}_1} & P_{\hat{m}\hat{y}_2} & \cdots \\ P_{\hat{y}_1\hat{x}_v} & P_{\hat{y}_1\hat{m}} & P_{\hat{y}_1\hat{y}_1} & P_{\hat{y}_1\hat{y}_2} & \cdots \\ P_{\hat{y}_2\hat{x}_v} & P_{\hat{y}_2\hat{m}} & P_{\hat{y}_2\hat{y}_1} & P_{\hat{y}_2\hat{y}_2} & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{2}$$

The camera's state vector $\hat{x}_v$ contains the position r^W, orientation R^{RW}, translational velocity v^W and angular velocity w^R of the camera. The periodic respiration model $\hat{m} = (\alpha, \tau, b, z_0)^T$ is represented by the parameters derived from Eq. 1, such that $z(t) = z_0 - b\cos^{2n}(\alpha)$ where $\alpha = \pi t / \tau$, t is the time step, z_0 is

the exhale position of the liver, b is the amplitude, τ is the frequency and $n = 3$ in accordance with [10]. Phase ϕ is not included as the system is initialized at $\phi = 0$.

The i^{th} map feature $\hat{y}_i = (\overline{y}, eig)$ is derived from the PCA transformation. $\overline{y} = (y_x, y_y, y_z)$ is the mean position of the feature in 3D space during a respiration cycle and $eig = (eig_x, eig_y, eig_z)$ is the eigenvector describing the transformation from 3D space to the periodic respiration model.

The **state prediction model** predicts camera, respiration and map motion. The camera motion is predicted using a "constant velocity, constant angular velocity" model. The state prediction model includes the addition step to predict the point in the respiration cycle and subsequently the motion in the map. The prediction model f_v^m and process noise covariance Q_v^m for the periodic respiration $\hat{m}$ are

$$f_v^m = \begin{bmatrix} 1 & t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad Q_v^m = \begin{bmatrix} \frac{\Phi_\tau t^3}{3} & \frac{\Phi_\tau t^2}{2} & 0 & 0 \\ \frac{\Phi_\tau t^2}{2} & \Phi_\tau t & 0 & 0 \\ 0 & 0 & \Phi_b t & 0 \\ 0 & 0 & 0 & \Phi_{z_0} t \end{bmatrix} \tag{3}$$

where Φ_τ is noise in the frequency, Φ_b is noise in the amplitude and Φ_{z_0} is noise in the exhale position. The predicted position of the i^{th} map feature in the world coordinate system y_i^W is computed using the predicted respiration parameters and $\hat{y}_i$ such that $y_i^W = eig(z_0 - b\cos^{2n}(\alpha)) + \overline{y}$.

The **measurement model** transforms the state space into the measurement space. Features in the map are measured relative to the camera. A feature's position in the camera coordinate system is estimated using $h_L^R = R^{RW}(y_i^W - r^W)$, where R^{RW} and r^W are the predicted camera rotation and position in the world coordinate system. y_i^W is the position of the feature in the world coordinate system as predicted by the respiration model. The measurement model is

$$h_L^R = R^{RW}(eig(z_0 - b\cos^{2n}(\alpha)) + \overline{y} - r^W) \tag{4}$$

and the partial derivatives (used in the EKF) of the measurement model with respect to $\hat{m}$ are

$$\frac{d\hat{m}}{d\alpha} = R^{RW} eig(nb\sin(\alpha)\cos(\alpha)^{n-1}), \qquad \frac{d\hat{m}}{d\tau} = 0 \tag{5}$$

$$\frac{d\hat{m}}{db} = -R^{RW} eig(\cos(\alpha)^n), \qquad \frac{d\hat{m}}{dz_0} = R^{RW} eig \tag{6}$$

The features are tracked in the image using [9]. During **system initialization** the camera is assumed static for one respiration cycle. Once initialized, new features can easily be added to the system when the camera is moving.

Table 1. Parameters for respiration modeling

	τ **(Frames)**	b **(cm)**	z_0 **(cm)**
Simulated Estimated	32.38	3.09	0. 95
Simulated Ground Truth	31.83	3	1
Ex Vivo Estimated	52.47	0.85	0.33
Ex Vivo Ground Truth	52	0.9	0.3

3 Results

The proposed MC-SLAM framework is validated on a simulated data set with known ground truth. It is also validated on an *ex vivo* tissue experiment with induced deformation and applied to *in vivo* footage. For quantitative validation with the simulated data, a virtual stereo laparoscope was navigated through a 3D virtual environment with periodic motion applied to a 3D mesh using Eq. 1 and the parameter settings shown in Table 1. The mesh was textured with an image of the liver.

Camera localization is evaluated in Figs. 3a-c) where MC-SLAM is compared to the ground truth and results from the static SLAM framework. The mean position error and standard deviation are 0.25cm and 0.19cm for MC-SLAM and 1.31cm and 0.6cm for static SLAM respectively. As the map and camera position are simultaneously estimated, accurate camera estimation is essential. The position error in MC-SLAM is attributed to rapid changes in acceleration of the camera's motion which are not well modeled. In this simulation, the dominant map motion is in the Z axis and it is evident that this map motion is absorbed into the static SLAM's estimation of the camera's position as there is a periodic error in the Z axis (Fig. 3c). Rotation is accurately recovered by both systems. To validate the method for modeling respiration, the estimated parameters are compared to the ground truth in Table 1.

For the *ex vivo* experiment performed, the ground truth position of the laparoscope was obtained using the approach in [11]. An *ex vivo* porcine liver sample was used and tissue motion was induced with a custom mechanical device. The device consisted of a cam, stepper motor and a sliding tray. Asymmetric motion is induced using Eq. 1 where z, n and b are defined by the cam profile and τ is defined by the stepper motor. Quantitative evaluation of laparoscope localization is shown in Fig 3d-f). The recovered motion using MC-SLAM closely follows the ground truth. Static SLAM however, periodically oscillates away from the ground truth. The mean error and standard deviation are 0.11cm and 0.07cm for MC-SLAM and 0.56cm and 0.25cm for static SLAM respectively. In addition, static SLAM is more prone to data association errors as shown in Fig. 3d-f) between frames 800-1000. The estimated and ground truth parameters of the respiration model are compared in Table 1 and Fig. 4a).

Figs. 4b-e) show the intra-operative laparoscopic images of the *ex vivo* tissue augmented with a virtual tumor which is manually and rigidly registered to the MC-SLAM map. The tumor is visualized using the Augmented Reality (AR) technique presented in [12]. Figs. 4b-c) are captured from a static laparoscope and illustrate the position of the tumor at full exhale and full inhale. Figs. 4d-e) demonstrate the system working in the presents laparoscopic and tissue motion. In Fig. 4e), the camera

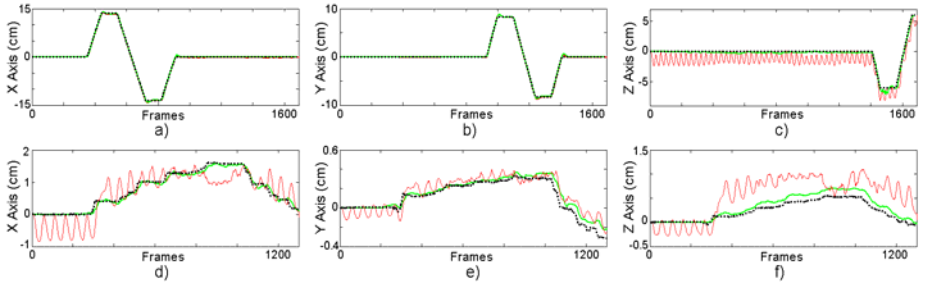


Fig. 3. Quantitative comparison of estimated laparoscope position in the world coordinate frame with MC-SLAM (green), static SLAM (red) and ground truth (black dashed). **(a-c)** simulated data XYZ axes. **(d-f)** *ex vivo* data XYZ axes.

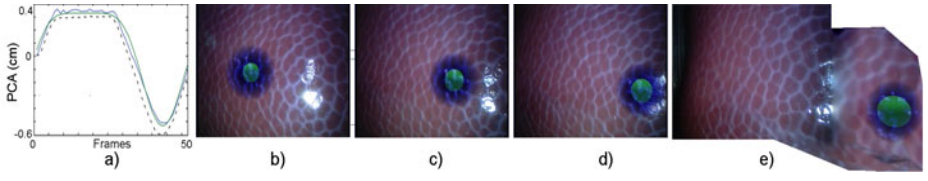


Fig. 4. *Ex vivo* experiment results, **(a)** the estimated respiration model; blue - observed data, green - respiration model, black dashed - ground truth. **(b-e)** Illustration of Image Guided Surgery with pre-operative data visualized [12] intra-operatively. **(a-b)** show a static laparoscope and the tissue at **(b)** exhale and **(c)** inhale position. **(d)** combined laparoscope and tissue motion. **(e)** laparoscope motion results in the target moving outside the current field-of-view. The dynamic target position is estimated relative to the current position of the laparoscope and visualized using view expansion [13].

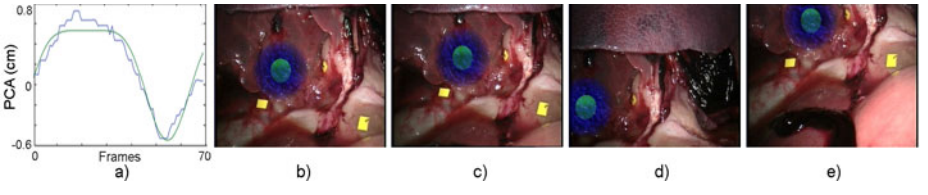


Fig. 5. *In vivo* experiment results; **(a)** the estimated respiration model (green) and observed data (blue); **(b-e)** illustration of Image Guided Surgery on *in vivo* footage with virtual pre-operative data visualized [12] intra-operatively; **(b-c)** images from a static laparoscope with the tissue at **(b)** exhale and **(c)** inhale position; **(d-e)** show combined tissue and laparoscope motion during abdominal exploration

navigates away from the tumor and its position is visualized outside the current field-of-view using dynamic view expansion [13]. This illustrates the capability of the system to predict the dynamic 3D position of tissue even when the tissue is not in the camera's current field-of-view.

For the *in vivo* experiment, the ground truth data was not available. The estimated respiration model is illustrated in Fig. 5a) and Figs. 5b-e) illustrate results with the use

of AR visualization. A virtual tumor is manually and rigidly registered to the MC-SLAM map. Figs. 5b-c) show intra-operative *in vivo* images captured from a static laparoscope. Figs. 5b) and 5c) show the tissue position at the full exhale and full inhale point in the respiration cycle. This illustrates tissue displacement resulting from respiration which was estimated at 1.08cm. The motion of the augmented tumor demonstrates the dynamic nature of the MC-SLAM map and progression beyond the rigid environment assumption. In Fig. 5d) and Fig. 5e), the laparoscope is navigated by the surgeon to explore the abdomen. Throughout the exploration the augmented data is displayed in a visually consistent manor in the presents of both laparoscope and tissue motion.

4 Conclusions

In this paper, we have presented a novel MC-SLAM system for simultaneous laparoscopic localization and dynamic tissue mapping. The system explicitly incorporates a periodic model of respiration into the statistical framework. This enables the system to predict and anticipate changes in tissue structure and estimated organ motion even when it is not in the laparoscope's current field-of-view. The method has been validated on simulated and *ex vivo* data and its clinical relevance has been demonstrated with *in vivo* data. Future work will focus on non-rigid registration of pre-operative data, faster initialization and more sophisticated motion models.

References

1. 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)
2. Garcıa, O., Civera, J., Gueme, A., Munoz, V., Montiel, J.M.M.: Real-time 3D Modeling from Endoscope Image Sequences. In: International Conference on Robotics and Automation Workshop on Advanced Sensing and Sensor Integration in Medical Robotics (2009)
3. Koppel, D., Chen, C.-I., Wang, Y.-F., Lee, H., Gu, J., Poirson, A., et al.: Toward automated model building from video in computer-assisted diagnoses in colonoscopy. In: Proc. SPIE (2007)
4. Wu, C.-H., Sun, Y.-N., Chang, C.-C.: Three-dimensional modeling from endoscopic video using geometric constraints via feature positioning. IEEE Transactions on Biomedical Engineering 54(7), 1199–1211 (2007)
5. Burschka, D., Li, M., Ishii, M., Taylor, R., Hager, G.D.: Scale-invariant Registration of Monocular Endoscopic Images to CT-scans for Sinus Surgery. Medical Image Analysis 9(5), 413–426 (2005)
6. Hu, M., Penney, G.P., Rueckert, D., Edwards, P.J., Bello, R., Casula, R., et al.: Non-rigid Reconstruction of the beating Heart Surface for Minimally Invasive Cardiac Surgery. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 34–42. Springer, Heidelberg (2009)

7. Stoyanov, D., Mylonas, G.P., Deligianni, F., Darzi, A., Yang, G.-Z.: Soft-tissue Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 139–146. Springer, Heidelberg (2005)
8. Davies, S.C., Hill, A.L., Holmes, R.B., Halliwell, M., Jackson, P.C.: Ultrasound quantitation of respiratory organ motion in the upper abdomen. British Journal of Radiology 67, 1096–1102 (1994)
9. Mountney, P., Yang, G.-Z.: Soft Tissue Tracking for Minimally Invasive Surgery: Learning Local Deformation Online. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 364–372. Springer, Heidelberg (2008)
10. Lujan, A.E., Larsen, E.W., Balter, J.M., Haken, R.K.T.: A Method for Incorporating Organ Motion due to Breathing into 3D Dose Calculations. Medical Phisics 26(5), 715–720 (1999)
11. Noonan, D., Mountney, P., Elson, D., Darzi, A., Yang, G.-Z.: A Stereoscopic Fibroscope for Camera Motion and 3D Depth Recovery During Minimally Invasive Surgery. In: International Conference on Robotics and Automation, pp. 4463–4468 (2009)
12. Lerotic, M., Chung, A.J., Mylonas, G., Yang, G.-Z.: pq -space Based Non-Photorealistic Rendering for Augmented Reality. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 102–109. Springer, Heidelberg (2007)
13. Mountney, P., Yang, G.-Z.: Dynamic View Expansion for Minimally Invasive Surgery using Simultaneous Localization And Mapping. In: Engineering in Medicine and Biology Conference, pp. 1184–1187 (2009)

Region Flow: A Multi-stage Method for Colonoscopy Tracking

Jianfei Liu[1], Kalpathi R. Subramanian[1], and Terry S. Yoo[2]

[1] Department of Computer Science,
The University of North Carolina at Charlotte, Charlotte, NC, 28223, USA
jliu1@uncc.edu, krs@uncc.edu
[2] Office of High Performance Computing and Communications,
National Library of Medicine, NIH, Bethesda, MD, 20892, USA
yoo@nlm.nih.gov

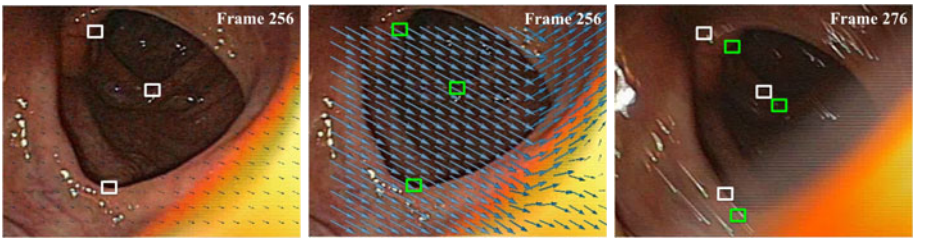
Abstract. Co-located optical and virtual colonoscopy images provide important clinical information during routine colonoscopy procedures. Tracking algorithms that rely on image features to align virtual and optical images can fail when they encounter blurry image sequences. This is a common occurrence in colonoscopy images, when the endoscope touches a wall or is immersed in fluid. We propose a *region-flow* based matching algorithm to determine the large changes between images that bridge such interruptions in the visual field. The region flow field is used as the means to limit the search space for computing corresponding feature points; a sequence of refining steps is performed to identify the most reliable and accurate feature point pairs. The feature point pairs are then used in a deformation based scheme to compute the final camera parameters. We have successfully tested this algorithm on four clinical colonoscopy image sequences containing anywhere from 9-57 consecutive blurry images. Two additional tabletop experiments were performed to quantitatively validate the algorithm: the endoscope was moved along a slightly curved path by 24 mm and along a straight path by 40 mm. Our method reported errors within 1-5% in these experiments.

Keywords: colonoscopy, tracking, failure recovery, image matching.

1 Introduction

The simultaneous use of pre-segmented virtual and optical colonoscopy images during routine endoscopic procedures provides useful clinical information to the gastroenterologist. Tracking algorithms must be employed to keep both image sequences aligned throughout the procedure. In medical images this presents a number of challenges: images can become blurry (endoscope touching a wall, fluid immersion), bright areas or tools may appear, etc. In these instances, endoscopic images over short periods of time may be devoid of features, causing tracking algorithms to fail. The goal of this work is to investigate new methods to skip such interruptions in the visual field and seamlessly continue tracking. Optical colonoscopy image sequences (our application of interest here) are particularly

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 505–513, 2010.



(a) (b) (c)

Fig. 1. Region flow vs. optical flow for describing large motion, (a) source image with overlaid optical flow vectors, (b) source image with overlaid region flow vectors, (c) target image after a 20 frame blurry sequence. White and green squares in the target image represent 3 selected regions in the image, and correspond to the white and green squares in the source images, after application of optical and region flow vectors. Region flow does a better job tracking the image motion. The lengths of the vectors in the source images represent the magnitude of the motion velocity.

challenging, due to deformation and other artifacts, and necessitate very robust algorithms.

The problem we study in this work can be stated as follows: given tracked images prior to a blurry image sequence, determine camera parameters (translation and rotation velocities) after the sequence, so as to continue tracking. There have been two general approaches to handling blurry images in endoscopic sequences: the use of a magnetic sensor[1,2] for tracking bronchoscopy images, and recovery from failures. It is unclear that this will work well with colonoscopy images, due to the more severe deformation effects, making sensor calibration a difficult task. An alternate approach is to use computer vision algorithms to locate and match corresponding features along the temporal dimension[3]. Termed wide-baseline matching[4,5], these methods find temporal correspondence through local comparison of feature descriptors, and results depend on the distinctness of the image features. Optical flow has also been used to track images[6], but typically does not work well for large changes between images. Fig. 1 illustrates an example, with Figs. 1(a),1(c) representing optical images bridging a blurry image sequence. The white squares in these images represent corresponding pairs generated by the optical flow field; they do not match up with the green squares, which roughly represent the positions of the true corresponding pairs.

In this work we present a computer vision algorithm to accurately match corresponding features representing large motion between images, to improve the robustness of tracking algorithms. Central to this approach is the use of *region flow*, a dense feature matching strategy that provides a basis and framework for understanding large motion. As indicated by the work of Brox et al.[7] and Liu et al.[8], dense feature correspondences can improve the accuracy of many vision applications, such as structure-from-motion, object recognition, and image retrieval. In the problem addressed here, region flow computation is a key and novel step of our method, for two reasons, (1) permitting the algorithm to limit the search space for accurately identifying corresponding features, and

(2) point-to-point correspondence relies on the intensity constancy model, which is generally not true for large motion; instead, invariance of a region's intensity distribution is a more reasonable assumption. The region flow matching step is followed by a sequence of refinements, that lead towards accurate computation of corresponding feature point pairs, and involves region-to-region and point-to-point matching steps, and false feature rejection. Finally, an image deformation based egomotion estimation method is used to recover camera parameters; we constrain the 2D-2D image deformation based egomotion determination problem[9] to a 3D-2D pose estimation problem by using depth values from a colon model. Fig. 2 details the various steps of our region flow based approach to handling large motion and thereby provide the means to skip blurry image sequences.

We demonstrate initial results using this method on four clinical colonoscopy sequences containing blurry images; two of these are in the sigmoid colon and the remaining two in the ascending colon. A table top validation experiment was also performed to quantify the accuracy of the method by acquiring two sequences in which the endoscope was moved 24 mm along a slightly curved path and 40 mm along a straight path.

2 Methods

The flow of our algorithm for recovering motion parameters after a blurry image sequence is described in Fig. 2. We describe in more detail the major steps of our method.

2.1 Region Flow Computation

Computing region flow is the key to efficiently determining feature point correspondences between images representing large motion.

Let $I_1(x, y)$ and $I_2(x, y)$ be a pair of normalized images, with $\overrightarrow{u} = (u_x, u_y)$ representing the region flow vector at point (x, y). The similarity between two regions of $I_1(x, y)$ and $I_2(x, y)$ with relative displacement $\overrightarrow{u}$ can be measured by Normalized Cross-Correlation(NCC) and given by

$$NCC(x, y, u_x, u_y) = \iint I_2(x + u_x, y + u_y) I_1(x, y) dx dy \tag{1}$$

Similar to optical flow computation[10], we use a global energy function to compute region flow, within a minimization framework.

$$E(u_x, u_y) = \iint \underbrace{\min(|1.0 - NCC(x, y, u_x, u_y)|, \alpha)}_{\text{Data constraint}} + \lambda \underbrace{\min((|\nabla u_x| + |\nabla u_y|), \beta)}_{\text{Smoothness constraint}} dx dy \tag{2}$$

where α and β are truncation values to prevent over-smoothing. λ is a parameter to balance data and smoothness constraints.

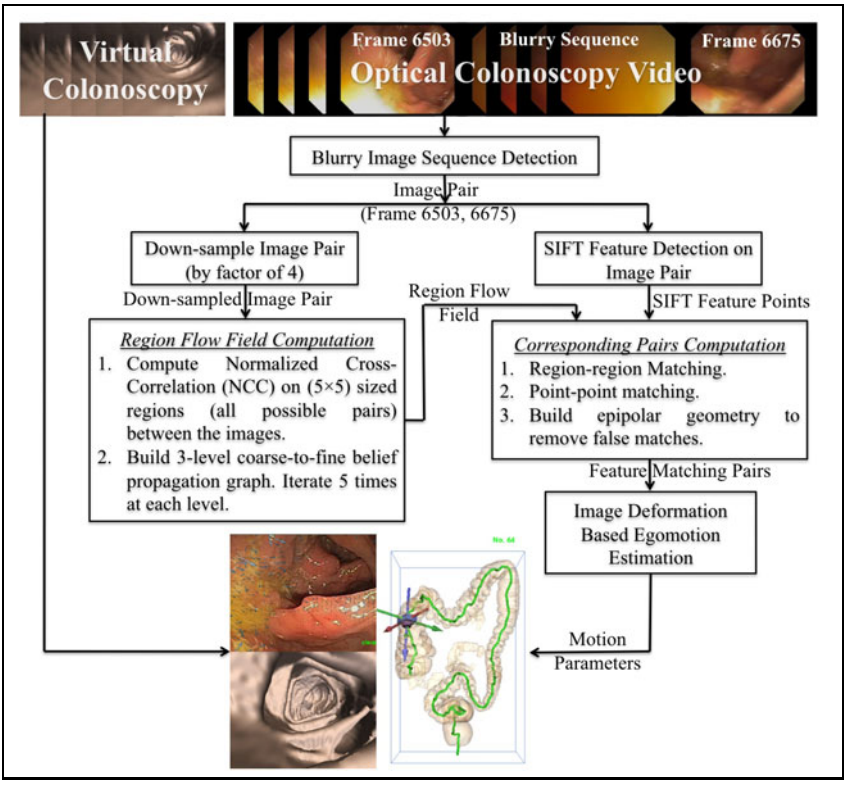
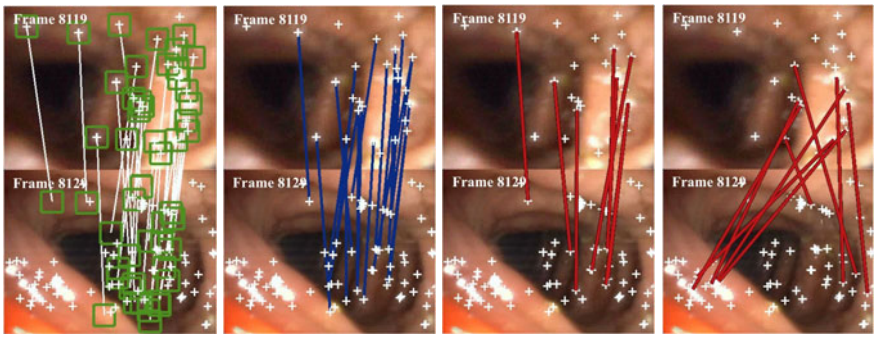


Fig. 2. Region flow based algorithm for recovering motion parameters after a blurry image sequence

Implementation. In general, region flow computation involves matching regions in the source image with regions in the target image, at every pixel, an $O(n^4)$ computation, for $N \times N$ sized images. In our implementation, we reduce the computation by (1) downsampling images by a factor of 4 in each dimension, and (2) restrict the largest image motion(we allow up to 150 pixel displacement on a 500 by 400 image), so that the corresponding search space in the target image is reduced. The minimization procedure of Eq. 2 is performed using the coarse-to-fine belief propagation procedure described in Felzenwalb[11]. The minimization results in a set of region flow vectors that provide a good approximation to the image motion. Fig. 1(b) illustrates an example colonoscopy image with overlaid region flow vectors. The region flow vectors follow the image motion between the two images, Figs. 1(b) and 1(c).

2.2 Corresponding Pairs Computation

Region-to-Region Matching. In this step, corresponding regions are identified using the region flow field and a local matching procedure. A set of stable feature points are detected by the SIFT algorithm[5] in the source and target



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

Fig. 3. Corresponding Pairs Computation. Top and bottom images represent images before and after the blurry image sequence, (a) Region-to-Region matching. Green squares indicate the matched regions using the region flow field. Local search using NCC is performed to find the best region pair (b) Point-to-Point feature matching. Using SIFT descriptor as a metric, the best SIFT feature point pair is determined between source and target regions. (c) False feature match rejection using epipolar geometry, (d) local matching using only (locally defined) SIFT feature descriptors, illustrating significant errors.

images. The corresponding regions in the target image are identified using the region flow vectors and a local neighborhood search. In Fig. 3(a), the green squares joined by the white lines represent corresponding regions containing at least one SIFT feature point in the source image and 0 or more SIFT feature points in the target region. In the implementation, the mapped region is locally adjusted using NCC as a metric to find the best region match.

Point-to-Point Feature Matching. In this step, each corresponding region pair is refined to a corresponding point pair. If the target region does not contain a SIFT feature point, it is removed. For target regions with multiple SIFT feature point candidates, the candidate with the closest SIFT descriptor(a distance metric) is chosen as the best candidate. Fig. 3(b) illustrates the selected feature point pairs after this step.

False Feature Match Rejection. With the chosen feature point pairs, epipolar geometry is built using the RANSAC algorithm[4]. Outliers that do not satisfy the epipolar geometry constraints are removed, as seen in Fig. 3(c).

Finally, Fig. 3(d) illustrates the same example using just SIFT feature point matching. It can be clearly seen that the lack of global motion information results in significant mismatches.

2.3 Image Deformation Based Egomotion Estimation

In the final step, we estimate the camera motion parameters of the image after the blurry images. We use a deformation based method.

In Fig. 4, the visual angle θ_1 of two world coordinate points P and Q from the camera projection center O_1 is defined as the angle between the projection rays of P and Q. If $\overrightarrow{v}_{o_1p_1}$ and $\overrightarrow{v}_{o_1q_1}$ are the normalized projection rays, the disparity between θ_1 and θ_2 can be expressed in terms of vectors orthogonal to $\overrightarrow{v}_{o_1p_1}$ and $\overrightarrow{v}_{o_1q_1}$ as follows(see [9]):

$$\theta_2 - \theta_1 = \overrightarrow{T} \cdot (dP\overrightarrow{v}^{\perp}_{o_1p_1} + dQ\overrightarrow{v}^{\perp}_{o_1q_1}) \qquad (3)$$

where $dP = \frac{1}{|P-O_1|}$, and $dQ = \frac{1}{|Q-O_1|}$ and $\overrightarrow{T} = (T_x, T_y, T_z)$ is the translation velocity. Eq. 3 depends only on $\overrightarrow{T}$. As depth values can be obtained from the virtual colon model, the computation of $\overrightarrow{T}$ is linearized. Once $\overrightarrow{T}$ is known, we can compute the *Focus of Expansion*, $FOE = (fT_x/T_z, fT_y/T_z)$. Rotation velocities $\overrightarrow{R} = (R_x, R_y, R_z)$ are then computed by embedding the FOE in a polar coordinate system, as described by Reiger [12].

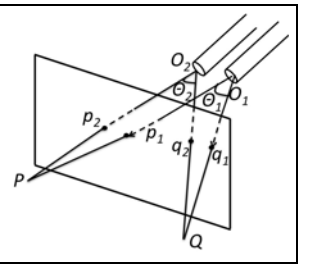


Fig. 4. Egomotion Estimation

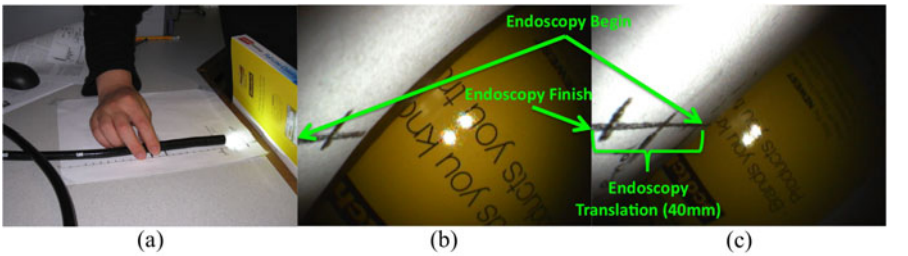


Fig. 5. Validation. A table top experiment was performed to move the endoscope a fixed distance. The resulting video was used to test the algorithm. (a) The table top experimental setup: the endoscope was placed on a surface with distance markings and moved along a predefined path, (b,c) Acquired images at the beginning and end of the sequence.

3 Experimental Results

We have tested our algorithm on four clinical colonoscopy sequences, distributed in different segments of the colon. We have also performed a validation experiment to quantify the accuracy of the algorithm.

3.1 Validation Experiment

Our goal in this experiment was to move the endoscope precisely a certain distance and acquire images between the end points. Since depth of the colonoscope from its starting point was needed, a flat surface was held at right angles at the edge of the desk, thus the depth is the same for all points in the projected image.

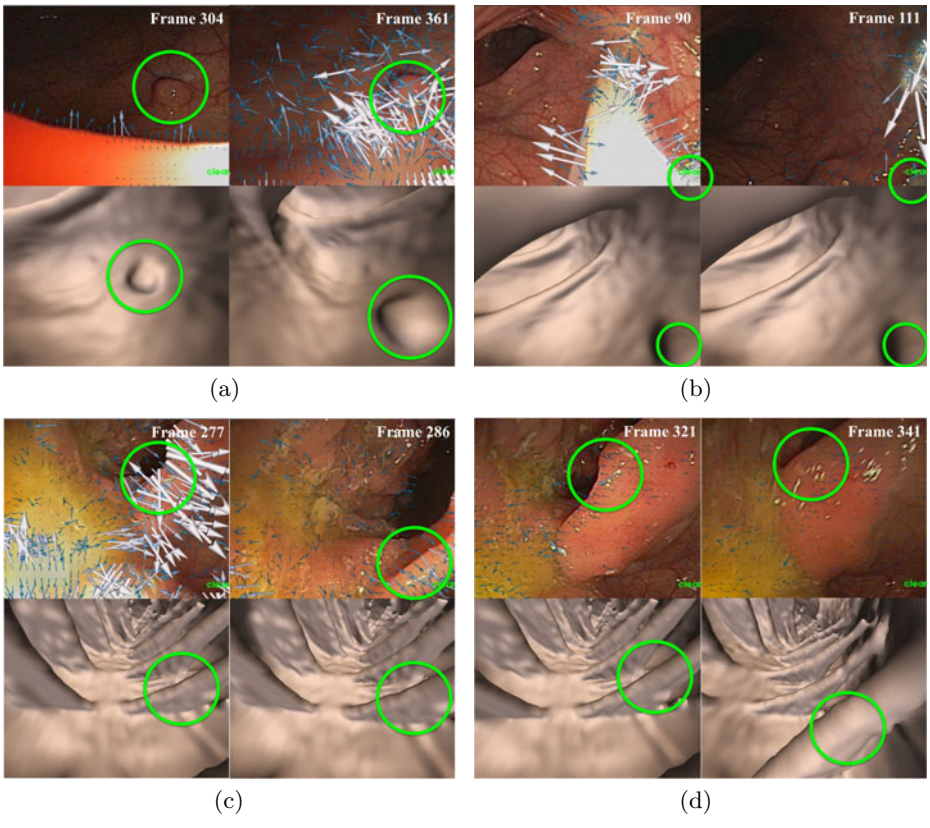


Fig. 6. Results on 4 Colonoscopy Sequences. OC-VC image pair before and after blurry sequences, (a) 520 polyp surgery sequence in sigmoid colon with a 57 image blurry sequence (b) 160 image polyp removal sequence in sigmoid colon with a 21 image blurry sequence, (c,d) 450 image sequence with 2 blurry image sequences of 9 and 19 images. The tracking system tracked through both blurry images sequences successfully.

Fig. 5 illustrates the experimental setup and image pair of one of the acquired sequences. Two sequences were tested with this setup, (1) colonoscope moved along a straight path by 40 mm, (2) colonoscope moved along a slightly curved path, with the end to end (Euclidean) distance of 24mm. Analyzing these two sequences using our algorithm resulted in distances of 39.6mm and 22.96mm respectively.

3.2 Clinical Colonoscopy Experiments

Fig. 6 illustrates 4 example colonoscopy sequences with blurry image sequences. The top rows illustrate the optical images before and after the blurry images. The corresponding virtual colonoscopy images are in the bottom rows. Regions marked by green circles indicate corresponding features to establish accuracy.

Experiment 1: Polyp Surgery in the Sigmoid Colon. This sequence contains 520 images, with a blurry image sequence from frame 304 to 361, due to the colonoscope touching the colon wall. In Fig. 6(a) the polyp can be clearly seen in the OC and VC images, including scale changes in the polyp. The fold in the virtual image is likely due to deformation.

Experiment 2: Polyp Removal in the Sigmoid Colon. This sequence represents the removal of the polyp, and contains 160 images, with a blurry image sequence between 90 and 111. Injection of water (bright area) in the vicinity of the removed polyp caused the blurry image sequence. Though somewhat harder to see, the green circles estimate the location of the polyp quite well in the OC and VC images.

Experiments 3,4: Ascending Colon. This sequence in the ascending colon contains 450 images and contained two blurry sequences, from 277 to 286 and 321-340; in both cases, the colonoscope was very close to a fold. Our algorithm was able to track continuously through the two blurry sequences, as seen by the well aligned OC and VC images in Figs. 6(c),6(d).

Our initial results are very promising. Despite the large changes in images in these sequences and the artifacts (especially deformation) in colonoscopy images, the region flow field accurately captures the global motion characteristics, easing the corresponding pairs computation. In all of these experiments, it is possible to identify features (folds, polyps, etc.) that provides qualitative accuracy and confidence in the tracking system. It is worthy to note that our tracking system tracked continuously through the two blurry image sequences in Figs. 6(c),6(d) without interruption.

4 Conclusions

We have presented a region flow based algorithm to handle large motion induced changes in colonoscopy video; this frequently happens when the colonoscope touches a wall or fold, or is immersed in fluid. The region flow field provides the computational basis for accurate and robust corresponding pairs computation, which in turn permits estimating camera parameters. We show through a validation experiment and four clinical colonoscopy sequences the effectiveness of our algorithm to keep the tracking system functioning as it encounters blurry image sequences; in our experiments, blurry image sequences ranged from 9 to 57 consecutive images. We are currently looking at two issues in improving our method, (1) carefully examine the computational considerations in computing region flow, and (2) improve the robustness of the egomotion estimation.

References

1. Mori, K., Deguchi, D., Akiyama, K., Kitasaka, T., Maurer Jr., C.R., Suenaga, Y., Takabatake, H., Mori, M., Natori, H.: Hybrid bronchoscope tracking using a magnetic tracking sensor and image registration. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 543–555. Springer, Heidelberg (2005)

2. Deligianni, F., Chung, A., Yang, G.Z.: Non-rigid 2d-3d registration with catheter tip EM tracking for patient specific bronchoscope simulation. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 281–288. Springer, Heidelberg (2006)
3. Sullivan, J., Carlsson, S.: Recognizing and tracking human action. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002. LNCS, vol. 2350, pp. 629–644. Springer, Heidelberg (2002)
4. Mikolajczyk, K., Schmid, C.: Scale and affine invariant interest point detectors. International Journal of Computer Vision 60(1), 63–86 (2004)
5. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
6. Liu, J., Subramanian, K., Yoo, T., Uitert, R.: A stable optic-flow based method for tracking colonoscopy images. In: Proc. of MMBIA, Anchorage, Alaska (2008)
7. Brox, T., Bregler, C., Malik, J.: Large displacement optical flow. In: Proc. of CVPR 2009, Miami, FL (June 2009)
8. Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T.: Sift flow: Dense correspondence across different scenes. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part III. LNCS, vol. 5304, pp. 28–42. Springer, Heidelberg (2008)
9. Zhang, J.: Computing Camera Heading: A Study. PhD thesis, Stanford University (1999)
10. Horn, B., Schunck, B.: Determining optical flow. Artif. Intell. 17(3), 185–203 (1981)
11. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. International Journal of Computer Vision 70(1), 41–54 (2006)
12. Reiger, J., Lawton, D.: Processing differential image motion. Journal of the Optical Society of America A 2(2), 354–359 (1985)

A System for Biopsy Site Re-targeting with Uncertainty in Gastroenterology and Oropharyngeal Examinations

Baptiste Allain[1], Mingxing Hu[1], Laurence B. Lovat[2], Richard J. Cook[3], Tom Vercauteren[4], Sebastien Ourselin[1], and David J. 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
[4] Mauna Kea Technologies, Paris, France

Abstract. Endoscopy guided probe-based optical biopsy is a new method for detecting sites for tissue biopsy. These sites need to be re-targeted precisely and accurately in subsequent images of the same endoscopy session for treatment or for visual guidance of surgical instruments. A new system for re-targeting biopsy sites and for characterising analytically their uncertainty is presented. It makes use of epipolar lines derived from multiple endoscopic views of the same site. It was tested on real patient data acquired during colonoscopy and gastroscopy. Gold standards of the biopsy site were obtained by Argon Plasma Coagulation tattooing. Re-targeting precision and accuracy were better than 0.8mm which is sufficient for most clinical endoscopic applications.

1 Introduction

Optical endoscopic techniques are being investigated for tissue interrogation *in vivo, in situ*. They aim to target biopsies better and should ultimately result in non-invasive replacements for traditional histological or cytological analyses of pathologies, in particular malignancies in accessible surfaces such as the epithelium of colon, oesophagus, or bladder [1]. A probe is often passed via an endoscope working channel and placed in contact with the tissue to interrogate a 0.5mm diameter site (**Fig. 1** a)). It is a significant challenge to accurately re-target within the same session a site examined by optical biopsy, in order to take a co-localised tissue biopsy (**Fig. 1** b), c)) [2, 3].

Navigation systems were developed for endoscope camera tracking in a pre-operative CT image and for instrument guidance [4, 5]. Large re-targeting errors may occur since the tissue in the organ being investigated is likely to deform significantly between the CT examination and the endoscopy. A simultaneous localisation and mapping system for the oesophagus could also solve the re-targeting [6]. This approach requires salient points that remain visible over the course of the examination for reliable results. This is difficult during endoscopy as only a small tissue extent is illuminated. In our paper, a biopsy site $\mathbf{p_{Ti}}$ whose position is known in a previous endoscopic image $\mathbf{I_i}$ is mapped onto a target image $\mathbf{T}$. This mapping is the fundamental matrix $\mathbf{F_{iT}}$. It is the composition of the camera's intrinsic parameters with its 3D

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 514–521, 2010.

motions [7]. A point in $\mathbf{I_i}$ is mapped onto $\mathbf{T}$ as an epipolar line, indicating the locus of possible matching points. An accurate line recovery requires the imaged tissue to be rigid or to only move rigidly in 3D space. Endoscopic images of the oesophagus typically show approximately 3cm x 3cm of tissue and this can be considered locally rigid [8]. Periodic tissue motions can be approximated by the affine transformation [8].

Re-targeting a biopsy site as the intersection of 2 epipolar lines was reported in [9]. We extend this method with epipolar lines derived from the locations of the biopsy site $\mathbf{p_{Ti}}$ in N different views $\mathbf{I_i}$ to improve accuracy and precision. We have implemented a robust system for biopsy site re-targeting in gastroenterology. It uses computer vision algorithms well-adapted to endoscopy [8, 10]. It measures analytically the re-targeting precision as an estimate of the accuracy. (i) First, we present a method for accurate and precise re-targeting of a potential tissue sampling point, which cannot be defined macroscopically in endoscopic images alone. (ii) Secondly, we present a method to characterise the re-targeted biopsy site uncertainty. Zhang defined the uncertainty of a vector as its covariance matrix and computed analytically the fundamental matrix uncertainty [11]. We applied this computation to the re-targeting problem. The analytical precision of the re-targeted biopsy site could be computed from its uncertainty in order to provide the endoscopist with a confidence measure. (iii) Finally, this system might be applied to any conventional or probe-based optical biopsy.

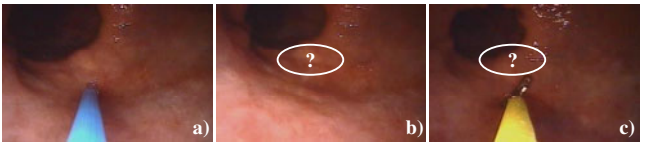


Fig. 1. Re-targeting problem: (a) endoscopic view with a probe; (b,c) forceps, for example, need to be guided to the biopsy site detected with the probe in order to take a biopsy

2 Method

2.1 Biopsy Site Re-targeting from Multiple Endoscopic Views

Consider two views $\mathbf{I_i}$ and $\mathbf{T}$ of the same region (**Fig. 2** a)). The epipolar line $\mathbf{el_i} = \mathbf{F_{iT}}\cdot\mathbf{p_{Ti}} = [el_{ix}, el_{iy}, el_{im}]^T$ is derived from the biopsy site $\mathbf{p_{Ti}} = [p_{Tix}, p_{Tiy}, m]^T$ expressed in homogeneous coordinates in $\mathbf{I_i}$ and from the fundamental matrix $\mathbf{F_{iT}}$. Ideally, this line passes through the true biopsy site and the epipole $\mathbf{e^{iT}}$. It is the intersection of the line joining the 2 camera centres with the image plane $\mathbf{T}$. Epipolar lines $\mathbf{el_i}$ derived from the images $\mathbf{I_i}$ have an associated error [7, 11], so they do not pass through the true biopsy site and they do not have a unique intersection. The site is estimated as the re-targeted biopsy site $\mathbf{p} = [p_x, p_y, m]^T$ that must satisfy the condition of triangulation with its match $\mathbf{p_{Ti}}$ [7]. This means that the two rays passing respectively through Camera centre i and $\mathbf{p_{Ti}}$, and through Camera centre T and $\mathbf{p}$, must meet at the position of the biopsy site $\mathbf{P}$ in the 3D space (**Fig. 2** a)). Two rays corresponding to the pair of matched points $\mathbf{p_{Ti}} \leftrightarrow \mathbf{p}$ will meet in space if and only if the points satisfy the familiar relationship termed algebraic distance [7]:

$$\mathbf{p}^{\mathrm{T}}\mathbf{F}_{\mathrm{iT}}\mathbf{p}_{\mathrm{Ti}} = \mathbf{p}^{\mathrm{T}}\mathbf{el}_{\mathrm{i}} = el_{ix}\cdot p_x + el_{iy}\cdot p_y + el_{im}\cdot m = 0\,. \tag{1}$$

As each pair of points $\mathbf{p_{Ti}} \leftrightarrow \mathbf{p}$ has to satisfy equation (**1**), the biopsy site **p** is computed as the point that minimises the sum of all the squares of the algebraic distances:

$$\min_{\mathbf{p}} \sum_{\mathbf{i}=1}^{N} C_i(\mathbf{el}_{\mathbf{i}}, \mathbf{p})^2 = \min_{\mathbf{p}} \sum_{\mathbf{i}=1}^{N} w_i^2 \left(\mathbf{p}^{\mathrm{T}}\mathbf{F}_{\mathrm{iT}}\mathbf{p}_{\mathrm{Ti}}\right)^2 = 0\,. \tag{2}$$

If the weights w_i are all set at 1, we minimise the algebraic distances that are not physically meaningful, while a measure in the 2D image plane **T** is a meaningful quantity [11]. It can be the distance from the biopsy site **p** to $\mathbf{el_i}$. Thus, w_i is determined as follows:

$$w_i = \frac{1}{\sqrt{el_{ix}^2 + el_{iy}^2}}\,. \tag{3}$$

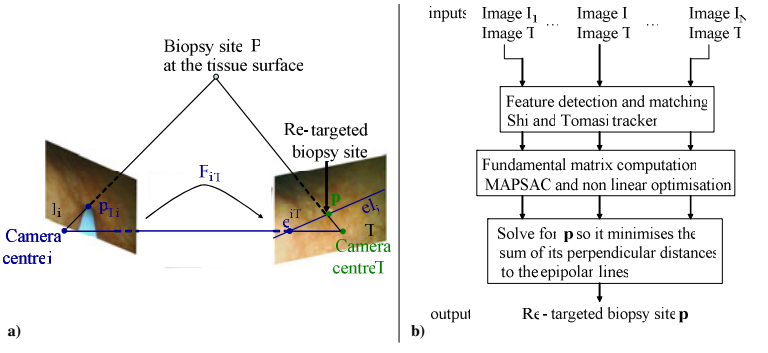


Fig. 2. Re-targeting principle: (a) the locus of the re-targeted biopsy site is indicated by the epipolar line $\mathbf{el_i}$ in **T**; (b) Framework of the re-targeting

2.2 Uncertainty and Analytical Precision of the Re-targeted Biopsy Site

The inaccuracy of the epipolar lines $\mathbf{el_i}$ cause uncertainty of the re-targeted biopsy site **p**. The covariance matrix of **p** noted $\Lambda_{\mathbf{p}} = E\left[(\mathbf{p}-E[\mathbf{p}])(\mathbf{p}-E[\mathbf{p}])^T\right] = \begin{bmatrix} var_{\mathbf{px}} & cov_{\mathbf{pxpy}} \\ cov_{\mathbf{pxpy}} & var_{\mathbf{py}} \end{bmatrix}$ models the uncertainty. From this matrix, a 95% confidence region around **p** can be displayed or the precision of **p** can be provided. Given $\Lambda_{\mathbf{p}}$, the precision of **p** is defined as:

$$\text{precision} = \sqrt{var_{\mathbf{px}} + var_{\mathbf{py}}}\,. \tag{4}$$

A general random vector $\mathbf{y} \in \mathbb{R}^k$ is handled instead of $\mathbf{p} \in \mathbb{R}^2$ for the derivation of Λ. The covariance matrix $\Lambda_{\mathbf{y(statistical)}}$ of **y** may be computed statistically from L samples $\mathbf{y}_i$ of **y**. Thus, the mean of **y** is $E_L[\mathbf{y}_i] = \frac{1}{L}\sum_{i=1}^{L}\mathbf{y}_i$ and $\Lambda_{\mathbf{y(statistical)}}$ is approximated by:

$$\Lambda_{\mathbf{y(statistical)}} = \frac{1}{L-1}\sum_{i=1}^{L}\left[(\mathbf{y}_i - E_L[\mathbf{y}_i])(\mathbf{y}_i - E_L[\mathbf{y}_i])^T\right]. \tag{5}$$

As the statistical method is time-consuming, Zhang [11] computed analytically $\mathbf{\Lambda}_{\mathbf{y(analytical)}}$ when $\mathbf{y}$ is measured from a random vector $\mathbf{x} \in \mathbb{R}^q$ so $\mathbf{y} = \varphi(\mathbf{x})$. This computation was applied to the fundamental matrix for uncertainty characterisation [11]. The matrix $\mathbf{\Lambda}_{\mathbf{y(analytical)}}$ is derived from $\mathbf{\Lambda}_{\mathbf{x}}$ and from the Jacobian matrix $\mathbf{D}_\varphi$ of φ as follows:

$$\mathbf{\Lambda}_{\mathbf{y(analytical)}} = \mathbf{D}_\varphi\left(E[\mathbf{x}]\right)\mathbf{\Lambda}_{\mathbf{x}}\mathbf{D}_\varphi\left(E[\mathbf{x}]\right)^T . \tag{6}$$

When $\mathbf{y}$ is recovered from $\mathbf{x}$ by the minimization of a criterion function, φ is unknown and $\mathbf{D}_\varphi$ is computed with the implicit function theorem, which states:

Proposition: *Let a criterion function C:* $\mathbb{R}^q \times \mathbb{R}^k \rightarrow \mathbb{R}$ *be a function of class* C^∞, $\mathbf{x_0} \in \mathbb{R}^q$ *the measurement vector and* $\mathbf{y_0} \in \mathbb{R}^k$ *a local minimum of* $C(\mathbf{x_0}, \mathbf{z})$. *If the Hessian* $\mathbf{H}$ *of C with respect to* $\mathbf{z}$ *is invertible at* $(\mathbf{x}, \mathbf{z}) = (\mathbf{x_0}, \mathbf{y_0})$ *then there exists an open set U' of* $\mathbb{R}^q$ *containing* $\mathbf{x_0}$ *and an open set U" such that for* $(\mathbf{x}, \mathbf{y})$ *in U'×U" there is an equivalence between the two relations '*$\mathbf{y}$ *is a local minimum of* $C(\mathbf{x}, \mathbf{z})$ *with respect to* $\mathbf{z}$*' and 'there exists a* C^1 *function* φ *such that* $\mathbf{y} = \varphi(\mathbf{x})$*', and:*

$$\mathbf{D}_\varphi(\mathbf{x}) = -\mathbf{H}^{-1}\frac{\partial \mathbf{\Phi}}{\partial \mathbf{x}} \text{ with } \mathbf{\Phi} = \left(\frac{\partial C}{\partial \mathbf{z}}\right)^{\mathrm{T}} \text{ and } \mathbf{H} = \frac{\partial \mathbf{\Phi}}{\partial \mathbf{z}} . \tag{7}$$

When C is of the form $C(\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{N} C_i(\mathbf{x}_i, \mathbf{z})^2$ and under the assumption that the noise in $\mathbf{x}_i$ and that in $\mathbf{x}_j$ (j≠i) are independent, the uncertainty $\mathbf{\Lambda}_{\mathbf{y(analytical)}}$ is:

$$\mathbf{\Lambda}_{\mathbf{y(analytical)}} = \frac{2 \cdot S}{N-k}\mathbf{H}^{-\mathrm{T}} \text{ with } \mathbf{H} \approx \mathbf{2} \cdot \sum_{i=1}^{N}\left(\frac{\partial C_i}{\partial \mathbf{z}}\right)^T \frac{\partial C_i}{\partial \mathbf{z}} . \tag{8}$$

where S is the value of the criterion C at the minimum.

In the case of the re-targeting of a biopsy site $\mathbf{p}$ with a series of N independent epipolar lines so $\mathbf{p}$ minimises the equation **(2)**, the derived criteria C_i have the form:

$$\left.\frac{\partial C_i}{\partial \mathbf{z}}\right|_{\mathbf{z}=\mathbf{p}} = \left[\frac{el_{ix}}{\sqrt{el_{ix}^2 + el_{iy}^2}}, \frac{el_{iy}}{\sqrt{el_{ix}^2 + el_{iy}^2}}\right] . \tag{9}$$

2.3 Re-targeting System

The epipolar lines $\mathbf{el_i}$ can be derived for the biopsy site re-targeting once the fundamental matrices $\mathbf{F_{iT}}$ have been computed for each pair of images $\mathbf{I_i} \leftrightarrow \mathbf{T}$ (**Fig. 2** b)).

Feature detection and matching: Images $\mathbf{I_i}$ and $\mathbf{T}$ need to be registered by detecting features in both images and by matching them. The endoscope acquires between $\mathbf{I_i}$ and $\mathbf{T}$ a series of intermediate images showing a probe in contact with the tissue and suddenly the tissue alone. As the displacement between two consecutive images is small, a feature detector and tracker through a video stream like the Shi and Tomasi tracker [12] is well-adapted. Features correspond to corners and are at the centre of template windows of $\mathbf{I_i}$. They are tracked between two consecutive images, assuming a small translation of the template. A refinement of the template position is performed between the first image $\mathbf{I_i}$, where the features were initially detected, and the current image, assuming an affine deformation.

Outliers' removal and computation of the fundamental matrix: The tracker may generate outliers that are wrong feature matches. They are detected using the Maximum A Posteriori Sample Consensus (MAPSAC) method [13]. This is a Monte Carlo-like process that draws samples of the matches to estimate the fundamental matrix $\mathbf{F}_{iT}$ and the inliers that return the best fit. Seven matches are drawn for each sample to estimate $\mathbf{F}_{iT}$ with the 7-point algorithm [7]. The result is refined by non-linear optimisation under the constraints that the rank of $\mathbf{F}_{iT}$ is 2 and its norm is equal to 1.

Biopsy site re-targeting and use of the uncertainty: A 95% error ellipse can be displayed around the re-targeted biopsy site and the precision can be returned.

3 Experiments and Results

3.1 Validation of the Analytical Precision by Simulations

Goal and method: The analytical precision (Eq. **4**, Eq. **8**) was compared to the statistical precision (Eq. **5**). Three-D points from an irregular semi-tubular surface (**Fig. 3** a)) were projected onto image planes $\mathbf{I_i}$ and $\mathbf{T}$ of various poses of a simulated camera. Eighty eight images were used to have the same number of images as for endoscopic sequences. They were 700 pixels x 700 pixels big as for endoscopic images. The projected points were the features. The biopsy site to re-target was selected from these.

A first experiment checked the goodness of the analytical precision in the presence of noise on the features. For each pair $\mathbf{I_i} \leftrightarrow \mathbf{T}$, 20% of the matches were displaced from 20 pixels to 40 pixels to create outliers. For a Gaussian noise of standard deviation varying from 0.5 pixels to 4 pixels and added to the inliers, the re-targeted biopsy site $\mathbf{p}$ and its analytical precision were estimated. This experiment was repeated 100 times to estimate, statistically, the covariance matrix and precision as in (Eq. **5**) and (Eq. **4**).

A second experiment checked the goodness of the analytical precision in the presence of outliers among the matches. A certain percentage of the matches were displaced as previously to create outliers in each pair $\mathbf{I_i} \leftrightarrow \mathbf{T}$, and a Gaussian noise of standard deviation 1.5 pixels was added to the inliers. The percentage of outliers was varied from 0% to 40%. The analytical precision was computed for each percentage and the statistical precision was computed from 100 repeats of the experiment.

Results and discussion: The analytical and statistical precisions had similar values (**Fig. 3** b) and c)). In addition, the re-targeted biopsy site tended to be less precise

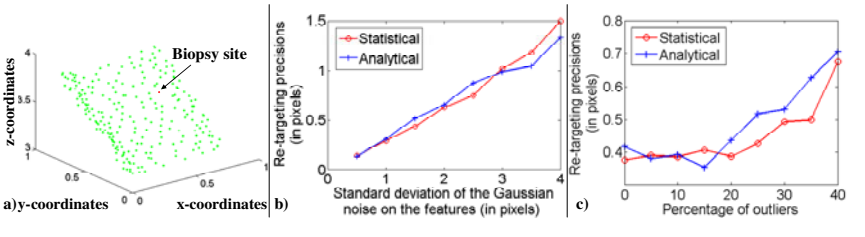


Fig. 3. a) Generated 3D points on an irregular semi-tubular surface; re-targeting precisions as functions of b) the noise on the features and of c) the percentage of outliers among the matches

when the noise and the percentage of outliers increased. Hartley and Zhang [7, 11] explained that the epipolar lines are determined with less accuracy in the presence of greater noise or with a higher percentage of outliers. Thus, they pass further away from the true biopsy site and their inaccuracy propagates to the re-targeted site.

3.2 Analytical Precision and Accuracy of the Re-targeting with Patient Data

Goal and method: The analytical precision and the accuracy of the re-targeting were estimated with real patient data from gastroscopy and colonoscopy. The data were acquired with both white-light and narrow band imaging wide-field endoscopes. Eight sequences were acquired on 4 patients during routine Barrett's Oesophagus surveillance procedures. Three other sequences were acquired on 2 patients during colonoscopy. For the gastroscopies, Argon Plasma Coagulation (APC) was used to create marks or gold standards of the biopsy site at the tissue surface. The re-targeting was tested on sequences showing the APC mark under various viewpoints (**Fig. 4**). For some sequences, images $\mathbf{I_i}$ showed the probe touching the mark and image **T** showed the mark alone (**Fig. 4**). For each sequence, the biopsy site position was tracked manually from $\mathbf{I_1}$ to **T** for the re-targeting in **T**. A region of interest around the mark in **T** was used as a mask to remove the corresponding features which would positively influence the accuracy of the fundamental matrix. As no mark was made in the colon, the biopsy site was selected as the probe tip or a vessel corner. The gold standard in **T** was given by interactive, manual tracking. The APC mark or the size of the probe provided a scale in **T** to convert the measures from pixels to millimetres. The precision was computed analytically. The accuracy of the re-targeting was computed in the 2D image **T** as the distance between the re-targeted biopsy site and the gold standard. This distance and the precision value are small enough to approximate the corresponding tissue extent as a plane. Under this assumption, the precision and the accuracy in millimetres in the 2D image are representative of the true measures in 3D.

Results and discussion: The accuracy and the analytical precision were 0.8mm or better (**Table 1**). The analytical precision was a reliable estimate of accuracy, showing that bias was minimal thus helping increase confidence in the re-targeting result.

The most precise and accurate results were obtained with sequences that were acquired without a probe in the FOV: robust features were detected with great disparity on the vessels and were tracked more easily. This guarantees accurate epipolar lines [7, 11]. In addition, the endoscope camera was moved freely and the resulting epipolar lines formed a bundle subtending a large angle. Their intersections were, therefore, close to the true position of the biopsy site and restricted the search space for the re-targeting. For Patient 4 Sequence 2 and Patient 5 Sequence 1, the images were more blurred than the other sequences, raising the possibility that the features were tracked with greater uncertainty and the epipolar lines were consequently less accurate. Only a small number of epipolar lines were used for the estimations, and they all subtended small angles to each other. Thus, their intersections could be further away from the true position of the biopsy site and the re-targeting was done with more uncertainty.

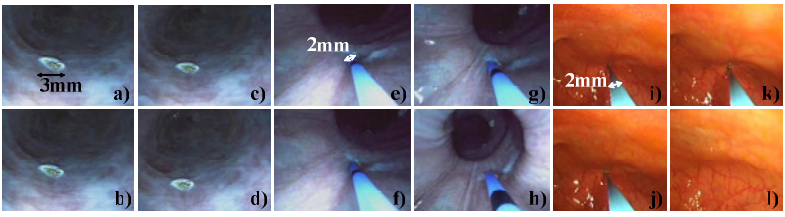


Fig. 4. Three endoscopic sequences: a), b), c), d) are images extracted from a gastroscopic sequence where the biopsy site (mark) was observed under various viewpoints; e), f), g), h) are from a gastroscopic sequence with a probe; and i), j), k), l) are from a colonoscopic sequence

Table 1. Precision (Eq. **4**) derived analytically from (Eq. **8**) and accuracy (distance between the gold standard and the re-targeted biopsy site) of the re-targeting in pixels and in millimetres

Sequence	Location	Probe?	FOV (pixels ; cm)	Precision (pixels ; mm)	Accuracy (pixels ; mm)
Patient 1 Sequence 1	Oesophagus	Yes	(339 x 216 ; 2 x 2)	(1.2 ; 0.12)	(1.3 ; 0.13)
Patient 1 Sequence 2	Oesophagus	No	(339 x 216 ; 5 x 2)	(0.3 ; 0.05)	(0.4 ; 0.08)
Patient 2 Sequence 1	Oesophagus	No	(283 x 180 ; 3 x 2)	(1.0 ; 0.05)	(1.3 ; 0.07)
Patient 3 Sequence 1	Oesophagus	No	(283 x 180 ; 6 x 2)	(0.6 ; 0.06)	(0.9 ; 0.09)
Patient 3 Sequence 2	Oesophagus	Yes	(283 x 180 ; 5 x 1)	(3.0 ; 0.20)	(6.3 ; 0.42)
Patient 4 Sequence 1	Oesophagus	Yes	(376 x 280 ; 1 x 1)	(2.6 ; 0.10)	(4.1 ; 0.16)
Patient 4 Sequence 2	Oesophagus	Yes	(376 x 280 ; 7 x 2)	(4.4 ; 0.44)	(4.5 ; 0.45)
Patient 4 Sequence 3	Oesophagus	Yes	(376 x 280 ; 1 x 1)	(15.0 ; 0.40)	(15.0 ; 0.40)
Patient 5 Sequence 1	Colon	Yes	(515 x 475 ; 3 x 3)	(5.1 ; 0.72)	(5.7 ; 0.82)
Patient 5 Sequence 2	Colon	No	(515 x 475 ; 3 x 3)	(0.9 ; 0.11)	(1.5 ; 0.18)
Patient 6 Sequence 1	Colon	No	(442 x 395 ; 3 x 3)	(1.5 ; 0.10)	(2.0 ; 0.13)

4 Conclusion

A biopsy site can be re-targeted in endoscopic images with N epipolar lines, derived from previous images where the biopsy site position is known and seen under different views. These views can be generated by twisting the head of the endoscope and by moving it gently up and down, while the probe remains at its location at the tissue surface. Sophisticated optics of endoscopes makes it easier to detect subtle features such as mucosal vascular patterns for epipolar line computation. This paper presents also a method for the analytical characterisation of the uncertainty of the re-targeted biopsy site to provide the endoscopist with a confidence measure, for example the precision, with which he/she can return to the original site. The system was tested on colonoscopy and gastroscopy data of patients. Analytical precisions and distances between the gold standard and the re-targeted biopsy site were better than 0.8mm.

Continued development will focus on a system of image-guided interventions. Such a system would detect microscopic pathologies by probe-based optical biopsy in combination with wide-field endoscopy and pre-procedure medical imaging (CT/MRI). It would also guide instruments to identified targets. To this end, the biopsy site indicated by the tip of the probe needs to be localised automatically. Moreover, as the endoscope

may have fast motions and occlusions in the field of view may occur due to air/water bubbles, the Shi and Tomasi tracker may fail to track the features. Thus, future work will need to focus on the feature detection and matching.

Acknowledgments. This work was done at UCLH/UCL with funding from the Department of Health, DLH Clinician Scientist Award to Dr. Cook, EPSRC (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 paper expresses the authors' views and not necessarily the funders'.

References

1. Wang, T.D., Van Dam, J.: Optical Biopsy: a New Frontier in Endoscopic Detection and Diagnosis. Clinical Gastroenterology and Hepatology 2(9), 744–753 (2004)
2. Dahr, A., Kristie, S.J., Novelli, M.R., et al.: Elastic Scattering Spectroscopy for the Diagnosis of Colonic Lesions: Initial Results of a Novel Optical Biopsy Technique. Gastrointestinal Endoscopy 63(2), 257–261 (2006)
3. Pohl, H., Roesch, T., Vieth, M., et al.: Miniprobe Confocal Laser Microscopy for the Detection of Invisible Neoplasia in Patients with Barrett's Oesophagus. In: Gut 57, pp. 1648–1653 (2008)
4. Mori, K., Deguchi, D., Ishitani, K., et al.: 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. 651–664. Springer, Heidelberg (2007)
5. Helferty, J.P., Sherbondy, A.J., Kiraly, A.P., Higgins, W.E.: Computer-based System for the Virtual Endoscopic Guidance of Bronchoscopy. Computer Vision and Image Understanding 108(1-2), 171–187 (2007)
6. Mountney, P., Giannarou, S., Elson, D., Yang, G.Z.: Optical Biopsy Mapping for Minimally Invasive Cancer Screening. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 483–490. Springer, Heidelberg (2009)
7. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge (2004)
8. Hu, M., Penney, G., Rueckert, D., et al.: 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)
9. Allain, B., Hu, M., Lovat, L.B., et al.: Biopsy Site Re-localisation Based on the Computation of Epipolar Lines from Two Previous Endoscopic Images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 491–498. Springer, Heidelberg (2009)
10. 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)
11. Zhang, Z.: Determining the Epipolar Geometry and its Uncertainty: A Review. International Journal of Computer Vision 27(2), 161–195 (1998)
12. Shi, J., Tomasi, C.: Good Features to Track, TR 93–1399, Cornell U (1993)
13. Torr, P.H.S.: Bayesian Model Estimation and Selection for Epipolar Geometry and Generic Manifold Fitting. International Journal of Computer Vision 50(1), 35–61 (2002)

Epitomized Summarization of Wireless Capsule Endoscopic Videos for Efficient Visualization

Xinqi Chu[1,5], Chee Khun Poh[1], Liyuan Li[1], Kap Luk Chan[2], Shuicheng Yan[3], Weijia Shen[1], That Mon Htwe[1], Jiang Liu[1], Joo Hwee Lim[1], Eng Hui Ong[4], and Khek Yu Ho[4]

[1] Institute for Infocomm Research, Singapore
{xchu,ckpoh,lyli}@i2r.a-star.edu.sg
[2] Nanyang Technological University, Singapore
[3] National University of Singapore, Singapore
[4] Dept of Gastroenterology & Hepatology, National University Hospital, Singapore
[5] University of Illinois at Urbana-Champaign, USA

Abstract. A video recording of an examination by Wireless Capsule Endoscopy (WCE) may typically contain more than 55,000 video frames, which makes the manual visual screening by an experienced gastroenterologist a highly time-consuming task. In this paper, we propose a novel method of epitomized summarization of WCE videos for efficient visualization to a gastroenterologist. For each short sequence of a WCE video, an epitomized frame is generated. New constraints are introduced into the epitome formulation to achieve the necessary visual quality for manual examination, and an EM algorithm for learning the epitome is derived. First, the local context weights are introduced to generate the epitomized frame. The epitomized frame preserves the appearance of all the input patches from the frames of the short sequence. Furthermore, by introducing spatial distributions for semantic interpretation of image patches in our epitome formulation, we show that it also provides a framework to facilitate the semantic description of visual features to generate organized visual summarization of WCE video, where the patches in different positions correspond to different semantic information. Our experiments on real WCE videos show that, using epitomized summarization, the number of frames have to be examined by the gastroenterologist can be reduced to less than one-tenth of the original frames in the video.

1 Introduction

In the early beginning of this century, Wireless Capsule Endoscopy (WCE) was introduced for the examination of the gastrointestinal tract, especially the small intestine where the conventional endoscopy is unable to reach [5]. Wireless capsule endoscopy is a non-invasive imaging technique. It has now changed the way doctors and clinicians performing the examination. The examination is no longer a real-time process as in traditional endoscopy. After the data recording (in the form of a video) is done by the capsule going through the entire gastrointestinal

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 522–529, 2010.

tract, clinicians have to sit down in front of a computer to review a video containing possibly more than 55,000 frames, and select the frames he or she considers important. This process is very time consuming, requiring full concentration of the gastroenterologists throughout.

Existing computational methods on WCE image processing focus on disease detection, e.g. bleeding detection [6, 7, 8, 10, 9, 11] so that gastroenterologists do not need to go through the entire video sequence. However, in actual clinical practice, the gastroenterologist would always like to confirm the detection results generated by the software and not taking any risk of missing something in the WCE examination. This motivates us to explore computational methods that can reduce the time spent in the examination by gastroenterologist. To our knowledge, up to now, the only relevant work to address this problem is by Iakovidis et. al [1]. They proposed an unsupervised summarization method of the WCE video by selecting the most representative images from the video. First, the frames in the whole video are clustered based on symmetric non-negative matrix factorization initialized by the fuzzy c-means algorithm. Then, one or a few representative frames are selected from each cluster. It can reduce the number of frames to be examined down to about 10% compared to the original video. However, such key frame extraction technique would inevitably miss some information in the 90% thrown away frames, even though they are similar to the representative frames in global appearance.

In this paper, we propose a novel method of epitomized summarization of the WCE video for efficient visual examination by gastroenterologist. The epitome model can generate a condensed summary of the original video. To ensure the necessary visual quality of the generated epitome for clinical examination, we introduce the constraint for local context preservation. We further introduce the spatial distributions for various semantic interpretations of the local features. Hence, we show that the epitome model can be used to generate semantically organized summarization which is helpful to let the doctor to naturally focus on the important visual information. The most significant benefits of the proposed method are that it produces a highly condensed summarization (less than 10% of original video volume) with almost no loss of visual information and provides a framework to flexibly organize the visual information in the summarization.

The rest of the paper is organized as follows. Section 2 describes the epitomized summarization model, *i.e.* the general formulation, the introduced constraints, and the derived learning algorithm. Section 3 presents the experimental results and quantitative evaluations on real data from the hospital. The conclusions are given in Section 4.

2 Epitomized Summarization

Epitome modeling

The epitome model seeks an optimally condensed appearance representation under which all the patches in the input image can be explained from the epitome [4]. It does not take into account the visual quality of the epitome. The

patches in the epitome may be distorted and artifacts may be introduced in order to explain different input patches. Also, important features with large local variances conveying higher degree of saliency of the local context may be smoothed out in the epitome. These adversary effects render certain areas in the epitome not recognizable to human, as shown in Fig. 1. In addition, the patches from the images could be placed anywhere in the epitome due to the flat prior of mapping. As a result, the epitome is too cluttered for visual examination by human. We introduce the constraints for local context preservation and semantic organization to generate the epitomized summarization of high visual quality for visualization.

To generate the epitomized summarization of a short sequence, we extend the image epitome [3] to learn the epitome from a set of N input image frames $\{\mathbf{I}_n\}_{n=1}^{N}$. The learning algorithm compiles a large number of patches drawn from these training images. Similar to [2, 3], the epitome model is formulated as a specified generative model.

Let $\{\mathbf{Z}_{nk}\}_{k=1}^{P}$ be a set of P patches from the image $\mathbf{I}_n$. Each patch $\mathbf{Z}_{nk}$ contains pixels from a subset of image coordinates S_{nk} in $\mathbf{I}_n$. For each patch $\mathbf{Z}_{nk}$, the generative model uses a set of hidden mapping $\mathcal{T}_{nk}$ from the epitome $\mathbf{e}$ to the coordinates $i \in S_{nk}$ in $\mathbf{I}_n$. Given the epitome $\mathbf{e} = (\boldsymbol{\mu}, \boldsymbol{\phi})$ and mapping $\mathcal{T}_{nk}$, a patch is generated by copying the corresponding pixels from the epitome mean and variance map,

$$p(\mathbf{Z}_{nk}|\mathcal{T}_{nk}, \mathbf{e}) = \prod\nolimits_{i \in S_{nk}} \mathcal{N}(z_{i,k,n}; \mu_{\mathcal{T}_{nk}(i)}, \phi_{\mathcal{T}_{nk}(i)}) \tag{1}$$

where coordinate i is defined on the input image. Under this generative model, it is assumed that each patch from every image is generated independently. The joint distribution is:

$$p(\{\{\mathbf{Z}_{nk}, \mathcal{T}_{nk}\}_{k=1}^{P}\}_{n=1}^{N}, \mathbf{e}) = p(\mathbf{e}) \prod_{n=1}^{N} \prod_{k=1}^{P} p(\mathcal{T}_{nk}) w_{nk} \prod_{i \in S_{nk}} \mathcal{N}(z_{i,k,n}; \mu_{\mathcal{T}_{nk}(i)}, \phi_{\mathcal{T}_{nk}(i)}) \tag{2}$$

where $p(\mathbf{e}) = const$, w_{nk} is introduced for local context preservation, and Gaussian-like spatial distributions for $p(\mathcal{T}_{nk})$ are used for semantic organization of visual features in the epitome.

Local context preservation

In many cases, distinctive local contexts contain significant local variation and often occupy small parts in the WCE images. To preserve the visual quality of distinctive local contexts in the epitomized summarization, a patch weight is introduced which favors the patch containing rich local context, i.e. having large local variance. The patch weight is defined as $w_{nk} = f(\sigma^2_{\mathbf{Z}_{nk}})$, where $f()$ is an increasing function of the variance. In this paper, a sigmoid function is used

$$w_{nk} = \left(1 + \exp\{-38(\sigma^2_{\mathbf{Z}_{nk}} - 0.1)\}\right)^{-1} + 0.003 \tag{3}$$

The effect of patch weight on the learning of the epitome from WCE images is shown in Fig. 1. Since the smooth parts of the normal tissues are abundant, more

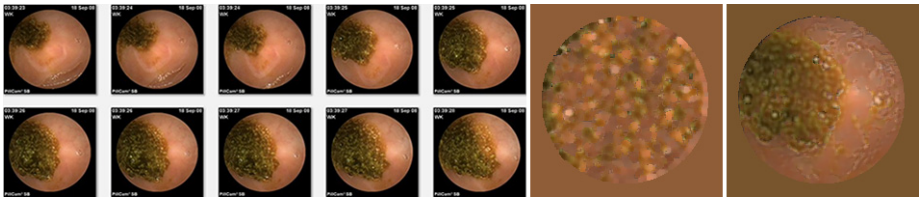


Fig. 1. The epitomes learnt from a sequence of 10 consecutive image frames. The left 10 images are the original frames with size of 288×288 pixels. The right two images are the enlarged epitome images. The epitome size is 200×200 pixels. The first epitome is generated by existing epitome formulation, and the second is the epitomized summarization generated by proposed method. The first epitome is too cluttered for human interpretation. In the second epitome, one can find the fine details in the regions of normal tissues from all the original frames.

details of rich context from all the 10 frames are learnt and placed in the region for normal tissues in the epitomized summarization. **Semantic Organization** A number of techniques for WCE image classification have been developed in the past decade. Even though the accuracy of the classification is not perfect, but if the semantic information of the classification can be integrated in the WCE summarization, it will be helpful for doctors to interpret the WCE videos.

In this work, we trained a Neural Network (NN) to classify each patch as one of three categories: Normal tissues, Non-informative contents (*i.e.* bubbles, fluids, and feces), and Suspected tissues (*i.e.* bleeding, tumor, lesion etc.). In a short sequence, if there are suspected tissues, we would like to place the related patches in the central region of the epitome summarization. The patches of normal tissues and non-informative contents are placed gradually further away from the central area in the epitome. In this way, if there are suspected tissues, it is easy to catch the attention of the gastroenterologist. This arrangement of patches in the summarized epitome according to the semantic attributes is naturally helpful to the gastroenterologist.

To implement the semantically organized epitome, we use three spatial distributions for the patches of the three semantic categories, as shown in Figure 2. For an input patch S_{nk} from one image frame of the short sequence, let $\mathcal{T}_{nk}(i)$ be a mapping from the epitome $\mathbf{e}$ to the image coordinate i, and $\mathcal{T}_0$ be the mapping to the center of the epitome. The prior of the mapping is defined as

$$p(\mathcal{T}_{nk}) = \begin{cases} \frac{1}{C_S} \exp\left\{-\frac{\|\mathcal{T}_{nk}(i)-\mathcal{T}_0\|^2}{\sigma_S^2}\right\}, & \text{if } S_{nk} \in \text{Suspected tissues} \\ \frac{1}{C_N} \exp\left\{-\frac{\|\mathcal{T}_{nk}(i)-\mathcal{T}_0\|^2}{\sigma_N^2}\right\}, & \text{if } S_{nk} \in \text{Normal tissues} \\ 1/|\mathbf{e}|, & \text{if } S_{nk} \in \text{Non-informative contents} \end{cases} \quad (4)$$

where $\sigma_S < \sigma_N$, $|\mathbf{e}|$ is the size of the epitome, and C_S and C_N are normalization constants.

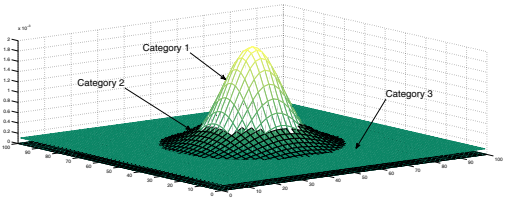


Fig. 2. The spatial distributions of the priors for patches of different semantic categories. The center one is for Suspected tissues, the surrounding one is for Normal tissues, and the flat one is for Non-informative contents.

EM learning

The epitomized summarization is generated by learning a generative model. Similar to [3], the variational inference is used to derive a new EM algorithm for the epitome formulation ([2]). The epitome's parameters and mapping distributions are estimated by optimizing the log likelihood of the observed patches using the approximate posterior to compute the lower bound on the log likelihood as in [3], we have the following updating equations: In the E-step, the posterior distribution over the hidden mappings $\mathcal{T}_{nk}$ is set to:

$$q(\mathcal{T}_{nk}) \sim p(\mathcal{T}_{nk}) w_{nk} \prod\nolimits_{i \in S_{nk}} \mathcal{N}(z_{i,n,k}; \hat{\mu}_{\mathcal{T}_{nk}(i)}, \hat{\phi}_{\mathcal{T}_{nk}(i)}) \tag{5}$$

In the M-step, from $\partial B/\partial \mu_j = 0$ and $\partial B/\partial \phi_j = 0$, the epitome mean $\hat{\mu}_j$ and variance $\hat{\phi}_j$ are updated as:

$$\hat{\mu}_j = \frac{\frac{1}{N}\sum_{n=1}^{N}\sum_{m=1}^{M} z_{nm} + \sum_{n=1}^{N}\sum_{k=1}^{P}\sum_{i \in S_{nk}}\sum_{\mathcal{T}_{nk}(i)=j} q(\mathcal{T}_{nk}) \frac{z_{i,n,k}}{2\hat{\phi}_j}}{M + \sum_{n=1}^{N}\sum_{k=1}^{P}\sum_{i \in S_{nk}}\sum_{\mathcal{T}_{nk}(i)=j} \frac{q(\mathcal{T}_{nk})}{2\hat{\phi}_j}} \tag{6}$$

$$\hat{\phi}_j = \frac{\sum_{n=1}^{N}\sum_{k=1}^{P}\sum_{i \in S_{nk}}\sum_{\mathcal{T}_{nk}, \mathcal{T}_{nk}(i)=j} q(\mathcal{T}_{nk})(z_{i,n,k} - \hat{\mu}_j)^2}{\sum_{n=1}^{N}\sum_{k=1}^{P}\sum_{i \in S_{nk}}\sum_{\mathcal{T}_{nk}, \mathcal{T}_{nk}(i)=j} q(\mathcal{T}_{nk})} \tag{7}$$

When the EM learning is complete, the epitomized summarization, *i.e.* the condensed representation of the short sequence, is generated with minimal loss of information.

3 Results

To illustrate and evaluate the performance of epitomized summarization of WCE videos, experiments were conducted on real WCE data from the hospital. In these experiments, two kinds of epitomized summarization were generated and evaluated. The difference between them is the computing of the prior distribution

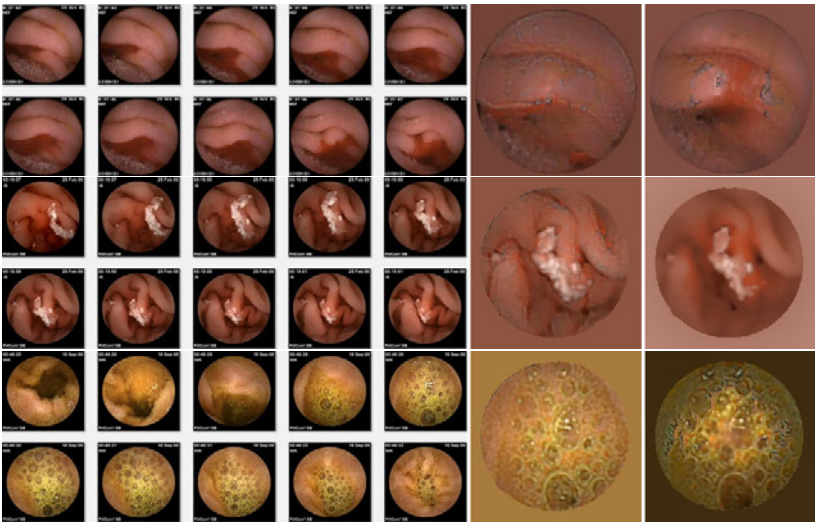


Fig. 3. Three more examples of learned epitomized summarization. From top to bottom, the examples show the cases of containing bleeding, bubbles, and substantial changes between consecutive frames. From the examples of bleedings, one can find the visual features of bleedings are concentrated and enhanced in the central regions in the semantically organized epitomes.

$p(\mathcal{T}_{nk})$. For the normal epitomized summarization, we randomly select a frame from the short sequence to initialize the epitome. That is equivalent to initialize the prior distribution $p(\mathcal{T}_{nk})$ with an image frame. So that learnt epitome summarization looks like the frames in the sequence, but not the same as any one of them. For the second type of epitomized summarization, the distributions for semantic descriptions given by (4) are used to generate the semantically organized epitome. When reviewing such epitomes, the doctors may naturally pay more attentions to the suspected tissues appearing in the central region and less attention to the non-informative contents appearing around the margins of the epitome image. Three more examples of normal epitomized summarization (NES) and semantically organized epitome summarization (SOES) are shown in Figure 3. For each row in the figures, the 10 images on the left (in two rows of

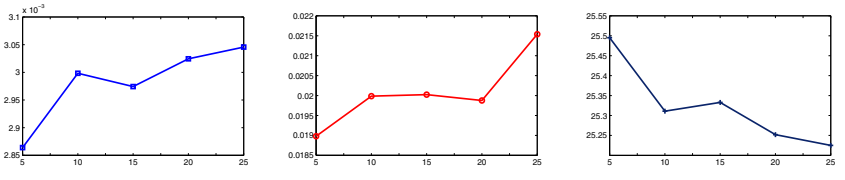


Fig. 4. The curves of the criteria over the length of the sequence N, where the left one is for $ASSD$, the middle one is for $MSSD$, and the right one is for $PSNR$

five columns) are the consecutive frames of the short sequence, and right-most two columns are the learnt NES and SOES. In these examples, the image size is 288×288 pixels, the epitome size is 200×200 pixels, and the patch size for epitome learning is 8×8 pixels. These examples show the visual quality of the epitomized summarization for the cases of bleedings, normal tissues, non-informative contents, and large motion between consecutive frames, etc. To quantitatively evaluate the visual quality of epitomized summarization, three criteria are used. First, for each patch $\mathbf{Z}_{nk}$ in the images of the sequence, we can find an epitome patch $\mathbf{e}_i$ which has the smallest difference with the image patch. The loss of the patch can be characterized as the normalized SSD (sum of squared difference):

$$SSD_{nk} = \min_{\mathbf{e}_i \in \mathbf{e}} \frac{1}{|\mathbf{e}_i|} \|\mathbf{Z}_{nk} - \mathbf{e}_i\|_2^2 \tag{8}$$

where $|\mathbf{e}_i|$ is the size of the patch. Based on this, three statistics can be computed as

$$ASSD = \frac{1}{NP} \sum_{n=1}^{N} \sum_{k=1}^{P} SSD_{nk} \tag{9}$$

$$MSSD = \max_{\forall n,k} \{SSD_{nk}\}_{k=1:P, n=1:N} \tag{10}$$

$$PSNR = 20 \log_{10} \left(\frac{MAX_I}{\sqrt{ASSD}} \right) \tag{11}$$

where $ASSD$ is the average of SSDs for all patches of the image frames from the sequence, $MSSD$ is the maximal SSD for all the patches of the images in the sequence which characterizes the maximum loss, and $PSNR$ is the peak signal-to-noise ratio.

We randomly select 50 sequences from the WCE videos of 7 patients, among them, half of the sequences contains various cases of bleedings. Each sequence maximally contains 30 consecutive frames, so that totally 1500 frames are used in the testing.

Let N be the length of the short sequences for summarization. Obviously, the larger the N is, the larger the reduction rate ($RR = N : 1$) is, but the loss of visual information will also increase (*i.e.* the $ASSD$ will increase and $PSNR$ will decrease). The loss of visual information with respect to the sequence length are evaluated. The curves of $ASSD$, $MSSD$ and $PSNR$ for N being 5, 10, 15, 20 and 25 are shown in Figure 4. It can be seen that there is no significant change of visual information loss even for $N = 25$. When $N = 15$ is used, i.e. at the point of $RR = 15 : 1$, the criteria values are: $ASSD = 0.00297$, $MSSD = 0.0200$, and $PSNR = 25.33$, which means the visual quality is acceptable for human examination [12]. The evaluation results indicate that, using epitomized summarization, it is possible to reduce the number of images down to less than 10% of the original videos with almost no loss of visual information for human examination.

4 Conclusion

We have proposed a novel approach of epitomized summarization of WCE videos to reduce the time spent on manual review of the recording by gastroenterologist. By introducing the constraints for local context preservation and semantic organization into the existing epitome framework, we show that the epitome technique can not only generate a highly condensed summarization with almost no loss of visual information, but also generated semantically organized visual summarization to naturally capture doctor's focus on relevant information. Quantitative evaluations have shown that it can reduce the frame number down to less than 10% with almost no loss of information. In our future work, we plan to extend the semantically organized epitome for WCE image registration [13].

References

[1] Iakovidis, D.K., Tsevas, S., Maroulis, D., Polydorou, A.: Unsupervised summarisation of capsule endoscopy video. In: 4th International IEEE Conference Intelligent Systems (2008)
[2] Cheung, V., Frey, B., Jojic, N.: Video epitomes. In: Proc. IEEE CVPR, pp. 42–49 (2005)
[3] Jojic, N., Frey, B., Kannan, A.: Epitomic analysis of appearance and shape. In: Proc. IEEE ICCV, pp. 34–41 (2003)
[4] Simakov, D., Caspi, Y., Shechtman, E., Irani, M.: Summarizing visual data using bidirectional similarity. In: Proc. IEEE CVPR, pp. 1–8 (2008)
[5] Iddan, G., Meron, G., Glukhovsky, A., Swain, P.: Wireless capsule endoscopy. Nature 405(6785), 417–418 (2000)
[6] Mackiewicz, M., Berens, J., Fisher, M., Bell, D.: Colour and texture based gastrointestinal tissue discrimination. In: Proc. IEEE Int'l Conf. on Acoustics, Speech and Signal Processing - Proceedings 2, ICASSP, pp. 11597–11600 (2006)
[7] Berens, J., Mackiewicz, M., Bell, D.: Stomach, intestine and colon tissue discriminators for wireless capsule endoscopy images. In: Proceedings of SPIE, Progress in Biomedical Optics and Imaging, vol. 5747, pp. (I): 283–290 (2005)
[8] Lee, J., Oh, J., Shah, S.K., Yuan, X., Tang, S.J.: Automatic classification of digestive organs in wireless capsule endoscopy videos. In: Proc. of the ACM, Symposium on Applied Computing, pp. 1041–1045 (2007)
[9] Bourbakis, N.: Detecting abnormal patterns in WCE images. In: Proc. 5th IEEE Symposium on Bioinformatics and Bioengineering, BIBE, pp. 232–238 (2005)
[10] Hwang, S., Oh, J., Cox, J., Tang, S.J., Tibbals, H.F.: Blood detection in wireless capsule endoscopy using expectation maximization clustering. In: Proceedings of SPIE, Progress in Biomedical Optics and Imaging, vol. 6144(I) (2006)
[11] Kodogiannis, V.S., Boulougoura, M.: Neural network-based approach for the classification of wireless-capsule endoscopic images. In: Proc. of the Int'l Joint Conference on Neural Networks, vol. 4, pp. 2423–2428 (2005)
[12] Wikipedia, Peak Signal to Noise Ratio, http://en.wikipedia.org/wiki/Peak_signal-to-noise_ratio
[13] Seshamani, S., Rajan, P., Kumar, R., Girgis, H., Dassopoulos, T., Mullin, G., Hager, G.: A Meta Registration Framework for Lesion Matching. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 582–589. Springer, Heidelberg (2009)

Computing Maximum Association Graph in Microscopic Nucleus Images*

Branislav Stojkovic[1], Yongding Zhu[1], Jinhui Xu[1],
Andrew Fritz[2], Michael J. Zeitz[3], Jaromira Vecerova[2], and Ronald Berezney[2]

[1] Department of Computer Science and Engineering,
State University of New York at Buffalo
{bs65,yzhu3,jinhui}@buffalo.edu
[2] Department of Biological Sciences, State University of New York at Buffalo
{ajfritz,vecerova,berezney}@buffalo.edu
[3] Stanford Medical School, 3801 Miranda Avenue, Palo Alto, CA 94304

Abstract. In this paper, we study the problem of finding organization patterns of chromosomes inside the cell nucleus from microscopic nucleus images. Emerging evidence from cell biology research suggests that global chromosome organization has a vital role in fundamental cell processes related to gene expression and regulation. To understand how chromosome territories are neighboring (or associated) to each other, in this paper we present a novel technique for computing a common association pattern, represented as a Maximum Association Graph (MAG), from the nucleus images of a population of cells. Our approach is based on an interesting integer linear programming formulation of the problem and utilizes inherent observations of the problem to yield optimal solutions. A two-stage technique is also introduced for producing near optimal approximations for large data sets.

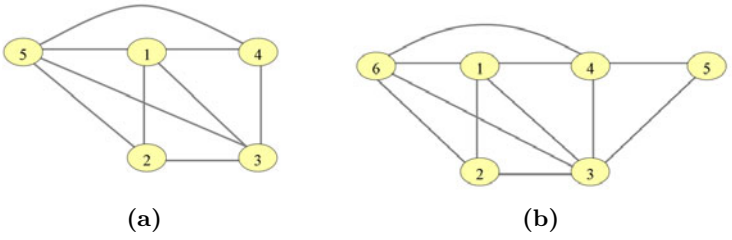
1 Introduction

Chromosome territory refers to a confined region within the cell nucleus where each chromosome is located. Empirical findings suggest that high level spatial organization of chromosome territories affects fundamental molecular biology processes [1,2,3]. In recent years, a significant amount of effort has been focused on studies of its influence on genomic structure and responsibility for specific nuclear functions, such as morphogenesis, transcription and splicing. Investigating association (spatial proximity) among chromosome territories is a necessary component for any study of the higher-order chromatin distribution effect on nuclear functions. However, despite recent developments in microscopy imaging and labeling techniques, accurately detecting such a pattern still remains a challenge.

To investigate this problem, one promising approach is to first represent the high level organization of the chromosome territories in each cell nucleus as a

* This research was partially supported by NSF through CAREER award CCF - 0546509, ROI grant IIS-0713489 and NIH grant GM-072131.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 530–537, 2010.



(a) **(b)**

Fig. 1. When finding a possible alignment from (a) to (b) GMG algorithm will choose the one with cost 4 (relabeling vertex 5 to 6, adding new vertex 5 and two incident edges). However, in matching where relabeling is disallowed, the only possible alignment is of cost 7 (removing two edges incident to vertex 5 and inserting vertex 6 with four incident edges).

graph, called *association graph*, and then find the common structural pattern from a set of association graphs corresponding to a population of cells. In each association graph, vertices represent the individual chromosome territories, and edges (between pairs of vertices) indicate the two corresponding chromosome territories are spatially proximate (or associated) to each other. Each vertex has a label which is its chromosome number. Since each chromosome has two homologs (or copies), a label is shared by a pair of vertices.

To find the common structural pattern from a set of association graphs, one way is to use the concept of graph median [4]. In the graph median problem, the input is a set of graphs with possible labels associated with nodes. The objective is to compute a new graph, called *median graph*, so that the distance between the median graph to the input graphs is minimized, where the distance between two graphs is normally defined as their edit distance. For the median graph problem, early results [4,5] are all of heuristic nature and do not have a bounded running time. Recently, a polynomial time combinatorial algorithm, called Generalized Median Graph(GMG) [6], was introduced by Mukherjee *et al.* Their algorithm uses a generalized distance function which allows vertex labels and edge weights, and produces near optimal solutions.

Despite its obvious advantages, the GMG (or any other median graph) algorithm also has two limitations when used for finding the association patterns from association graphs. First, due to its emphasis on finding the structural similarity between graphs, the GMG tends to match vertices with similar degrees. A possible outcome is that the GMG could match two vertices with different labels. While this is in accordance with the edit distance definition, it gives misleading semantic interpretation in our application to nuclear images, since mismatching has little biological meaning (see Figure 1). Second, the GMG requires graphs with unique predefined labeling. However, in association graphs, each label is shared by a pair of vertices. This is because in nuclear images each chromosome (including its two copies) is uniquely identified by its color, see Fig.2. Thus we are facing the challenge of how to correlate labels of chromosome copies among different cells. Even in the cases when professional expertise is at hand, it is not

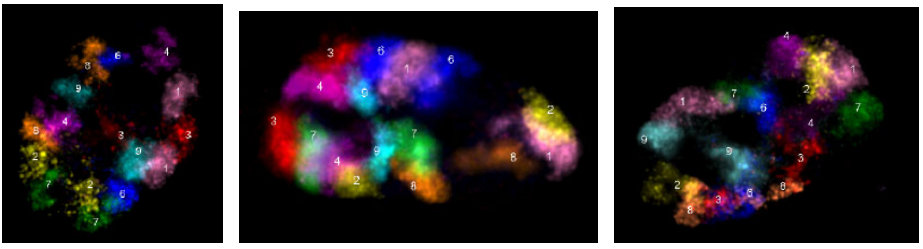


Fig. 2. Three different cell images (from the input set), obtained by superimposing z-stack images, containing eight chromosome colors(labels)

always possible to come up with a reliable mechanism for mapping copies of the same chromosomes between different cells.

To overcome all these difficulties, we propose in this paper a novel technique called *Maximum Association Graph*. Our approach maximizes the structural similarity between the computed new maximum association graph (MAG) and the set of input association graphs. During the matching process, it explicitly disallows matching of the vertices with different labels, and automatically establishes the correspondence of the same copy of the chromosome over all cells, thus eliminating the possible human errors. Also, it allows the user to specify a frequency threshold for determining the existence of an edge in the association pattern MAG (i.e., an edge in MAG if and only if it appears in a certain percentage of the input graphs). With this threshold, we could obtain the association patterns at different similarity levels.

Our approach is based on an mixed integer linear programming (MILP) formulation, and can be solved optimally using an optimization software package CPLEX for middle sized data sets. For large data sets, we also propose a two-stage approximation technique for obtaining near optimal polynomial-time solution.

2 Method

2.1 Maximal Structural Matching

Let L_V be a finite set of node labels and $\mathbb{I}_L \subseteq \mathbb{N}$ be the set of indices. A labeled undirected graph G, in which existence of more than one vertex with the same label is permitted, is defined as a triple $G \equiv (V, E, f^v)$ where V is the vertex set, E is the edge set, and $f^v : V \longrightarrow (L_V, \mathbb{I}_L)$ is a mapping which to every vertex v in V assigns unique pair (l, k), where $l \in L_V$ and $k \in \mathbb{I}_L$. We call node v a k-th instance (copy) of label l in G. In this way, each edge $e \in E$ between v_1 and v_2 is uniquely identified with 4-tuple (l_1, k_1, l_2, k_2), where $f(v_1) = (l_1, k_1)$ and $f(v_2) = (l_2, k_2)$.

Let $\mathbb{G} = \{G_1, G_2, \ldots, G_n\}$ be a collection of graphs where for every $G_i = (V_i, E_i, f_i^v)$, the following conditions hold: $V_i \subseteq V$ and $E_i \subseteq E$, and there is

no restriction on the cardinality of nodes in the graphs in G, i.e., $|V_i| \neq |V_j|$, $G_i, G_j \in \mathbb{G}$.

Structural pattern for $\mathbb{G}$ is obtained as maximal association graph, $\hat{G} = (\hat{V}, \hat{E}, f^v)$, where $\hat{V} = \bigcup_i V_i$ and $\hat{E} \subseteq \hat{V} \times \hat{V}$. As noted before, finding such a graph, $\hat{G}$, identifies the maximal set of edges between nodes in $\hat{V}$ that are consistent with labeling in all input graphs. This induces label preserving mapping from vertices in every input graph G_i to the set of vertices in $\hat{G}$. Which mapping is valid for particular node in some input graph is not known in advance, hence to obtain the optimal solution, we must look for all mapping simultaneously. This condition leads to the choice of mathematical optimization as a natural problem modeling tool. Although technique can be easily extended to deal with more general situation, we restrict our presentation to the case when each label has exactly two copies in every graph.

2.2 Mixed Integer Linear Programming Formulation

In our approach, we actively try to match vertices between every input graph and the resulting $\hat{G}$. A matching of an edges follows implicitly. Since, only matching of nodes with the same label is allowed, each node v in i-th graph representing k-th instance of label j, denoted as $v = (j, k)$, $v \in V_i$, can be matched to $\hat{k_1}$-th or $\hat{k_2}$-th instance of j-th label in graph $\hat{G}$. These two possible matchings have corresponding binary variables: $\alpha_{i,j,k,\hat{k_1}}$ and $\alpha_{i,j,k,\hat{k_2}}$. Value of one denotes that particular matching holds in the global solution. Similarly, for each node in $\mathbb{G}$, we introduce corresponding variables for all possible assignments. Thus, the set of constraints guaranteeing that each node in $\mathbb{G}$ has one and only one corresponding node in $\hat{G}$ is given as:

$$\sum_{k} \alpha_{i,j,k,\hat{k}} = 1, \forall i, j, \hat{k} \qquad (1)$$
$$\sum_{\hat{k}} \alpha_{i,j,k,\hat{k}} = 1, \forall i, j, k$$

For each potential edge in $\hat{E}$ between instance $\hat{k_1}$ of vertex labeled $\hat{j_1}$ and instance $\hat{i_2}$ of vertex labeled $\hat{j_2}$ there is a corresponding binary variable $\hat{\gamma}_{\hat{j_1},\hat{j_2},\hat{k_1},\hat{i_2}}$, which has the value of 1 if there is an edge between vertices $(\hat{j_1}, \hat{k_1})$ and $(\hat{j_2}, \hat{k_2})$ in $\hat{G}$, and 0 otherwise.

For each induced matching of an edge between instance k_1 of vertex labeled j_1 and instance i_2 of vertex labeled j_2 in G_i to an edge between instances $\hat{k_1}$ and $\hat{k_1}$ of corresponding vertex labels in $\hat{G}$ there is an associated real variable $\beta_{i,j_1,j_2,k_1,k_2,\hat{k_1},\hat{k_2}}$.

Formally, we denote an existence of an edge between vertices (j_1, k_1) and (j_2, k_2) in the input graph G_i with $\gamma_{i,j_1,j_2,k_1,k_2}$. Now, we can formulate matching constraint for such an edge as:

$$\beta_{i,j_1,j_2,k_1,k_2,\hat{k_1},\hat{k_2}} = \alpha_{i,j_1,k_1,\hat{k_1}} * \alpha_{i,j_2,k_2,\hat{k_2}} * \hat{\gamma}_{\hat{j_1},\hat{j_2},\hat{k_1},\hat{k_2}} * \gamma_{i,j_1,j_2,k_1,k_2} \qquad (2)$$

In other words, an existing edge (j_1, j_2, k_1, k_2) in input graph G_i can be matched to an existing edge $(\hat{j_1}, \hat{j_2}, \hat{k_1}, \hat{k_2})$ in $\hat{G}$ if and only if its belonging vertices (j_1, k_1) and (j_2, k_2) are matched to $(\hat{j_1}, \hat{k_1})$ and $(\hat{j_2}, \hat{k_2})$ in $\hat{V}$. Constraints given in (2) are in the nonlinear form, however using standard strategy of product term linearization in a constraint we can transform each constraint of this type to a set of inequality constraints.

For an edge in $\hat{G}$ to be present, its matching defined analogue must appear in at least a predefined percentage P of input graphs. This is ensured with the following set of constraint:

$$\sum_i \beta_{i,j_1,j_2,k_1,k_2,\hat{k_1},\hat{k_2}} \geq \hat{\gamma}_{\hat{j_1},\hat{j_2},\hat{k_1},\hat{k_2}} * P * |\mathbb{G}| \tag{3}$$

where P denotes the specified frequency threshold and $|\mathbb{G}|$ is the cardinality of input set.

Setting the expression for optimization function as the maximization of both number of edges in resulting graph as well as number of possible edge matchings from the set of input graphs we obtain the final mixed integer programming form.

2.3 Extensions to the Method

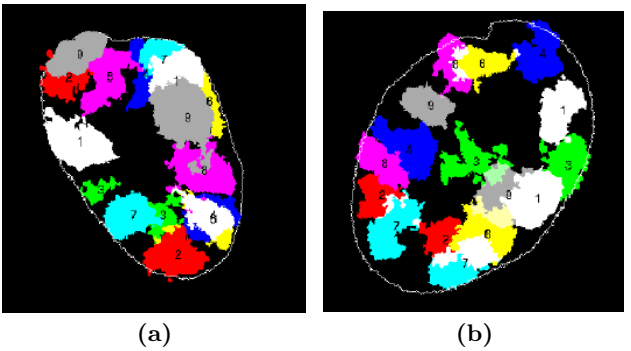
Including Pairwise solutions. In a restricted case we consider a pair of node labels and their belonging copies. Subgraphs obtained by isolating them from all other node instances in each graph from the input set $\mathbb{G}$ are passed as an input to algorithm, eq. (1-3). Solution for this subproblem provides the upper bound on the number of associations between four nodes in $\hat{G}$, two for each label from the chosen pair. In the first phase we solve MILP for each of $\binom{|L_V|}{2}$ pairs in order to obtain the upper bound for each of $\hat{\gamma}_{\hat{j_1},\hat{j_2},\hat{k_1},\hat{k_2}}$, $\hat{G}$ edge indicator variable. In the second phase we find the MILP solution, eq. (1-3), for original input set, but this time, adding set of upper bound constraints obtained in the first phase. Introducing new constraints is meaningful since it reduces feasible space and decreases overall running time without affecting the quality of the solution.

Two-stage algorithm. Mixed integer linear programming proposed above, eq. (1-3), can be solved optimally applying standard techniques of branch and cut available in commercial or free solvers. However, when dealing with large graphs or with increasing the size of input sets running time may considerably increase. Therefore we develop a modification for trading a little of accuracy for great reduction in the running time. We divide input sets in subgroups, and then solve problem on each of them optimally. We merge subgroup solutions by solving the MILP, eq. (1-3), again but this time on subgroup solutions given as input. We assign a proper weight to every edge denoting the number of graphs in which it appears. Exhaustive enumeration analysis suggests that a solution obtained in this way is close to the one obtained by solving MILP without partitioning. Over 95% of edges were preserved in approximative solution comparing to the optimal one.

3 Experimental Results

To test the validity of the method we calculated a maximum association graph on the input set of 15 randomly generated graphs. Graphs were created in the form of adjacency matrices mimicking the sparsity property of graphs obtained from cell nucleus images by having roughly the same number of edges. Adjacency matrices had at most 16 vertices with each vertex label appearing at most twice, which correlates with our real data graph size. A random graph A and then its complement $\tilde{A}$ were generated. We populated the first set with $P * 15$ number of permuted versions of graph A, and the second set with $(100\% - P) * 15$ number of permuted versions of graph $\tilde{A}$. Merging two sets together we obtained the input set. Using **CPLEX** as MILP solving tool, we were able to get the optimal solution i.e. calculated maximum association graph was identical to the generated input graph A in all test cases.

Further, the validation process considered graphs with added *noise*. We designed a procedure similar to those proposed in [7,8]. Following the steps from the previous paragraph a random graph A was created. We used graph A as the basis for populating the input set: For each new member of input set a copy of A, A', was made and randomly permuted with label preserving permutation matrix. To introduce the noise, removal of each edge in A' was considered with the probability denoted as noise parameter, E. In this way we ensured that all generated graphs were proper subgraphs of A having the expected number of associations equal to the value of noise parameter, E. Maximum association graph was compared with each graph in the input set individually. Cardinality of maximum matching was calculated (using permutation matrices) and normalized by the size of maximum possible matching for the whole input set (obtained by matching with the base graph A). Average values for two methods are shown in Table 1. They correlate (are slightly above or below) to the values of expected maximum number of associations that are allowed by introduced random perturbations. Also, we notice the agreement between the results of two proposed methods.



(a) **(b)**

Fig. 3. 2D segmented sections of chromosome territory in cell nucleus images

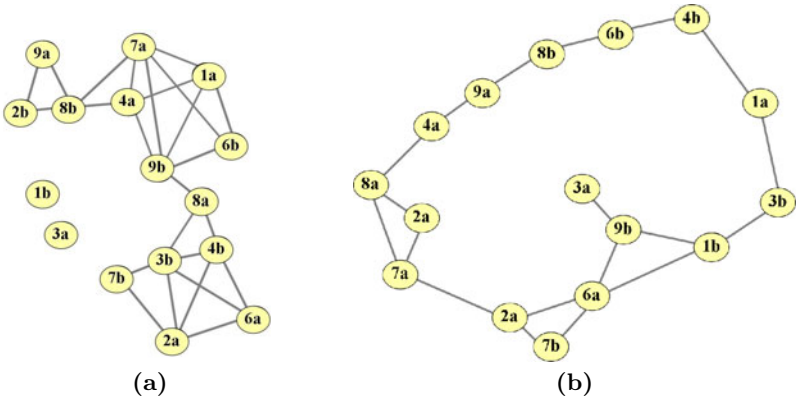


Fig. 4. Derived graphical representation of chromosome territory in cell nucleus

For qualitative evaluation of the proposed method we discuss experiments related to our motivating biological image analysis application. A previously developed procedure [9] using so called FISH technique was used for labeling 8 pairs of the larger chromosomes (numbers 1, 2, 3, 4, 6, 7, 8, 9) from the human lung fibroblast cell line WI38. The 3D microscopy z-stack image acquisition process was repeated for a total number of 45 cells, see Fig.2. After the registration and segmentation processes were performed, masks for individual chromosome pairs were created for each raw image, see Fig.3. Nearest border to border distances were calculated for every combination of chromosome pairs. To obtain the individual graph representation for each cell, pairwise associations were derived using a threshold value for distance of 4 pixels, which corresponds to 0.28 μm, see Fig.4. In the graph, each individual chromosome is represented as a labeled vertex. We distinguish between two instances of the same label by denoting them as copy a and copy b, respectively. Spatial proximity (adjacency) between chromosomes is expressed as an edge between vertices.

We report on the following observations. MAG results for different values of frequency threshold ($100\% - 40\%$) indicate consistency among the structural patterns. For example, the majority of the edges that appear at $P = 60\%$ also

Table 1. Average matching similarity with a graph from the input set

	Percentage of Matching		
Noise parameter (E)	15%	30%	40%
MAG	0.84	0.72	0.62
Two-Stage Method	0.83	0.69	0.60

Table 2. Average Sorensen's index value with a graph from the input set

Algorithm	MAG 65%	MAG 55%	MAG 50%	GMG
Sorensen Index	0.28	0.37	0.46	.40

appear at lower values for the threshold and we can almost perfectly align output graph for higher P as a sub-graph of computed maximum associated graph for lower P. Moreover, the model captures the essential association information from the constituent graphs. This is reflected in calculated Sorensen index values which show increased similarity with the decrease of frequency parameter P. Table 2 contains obtained values and provides comparison with the GMG method [9,6].

4 Conclusions

In this paper we used the specific form of mathematical optimization approach for computing the maximum association graph in context of a real world vision problems in biological image analysis. This formulation can be easily extended to handle the more general case, where arbitrary number of instances of vertices with the same label are present. This is relevant for the study of cancer cells where more than two copies of certain chromosomes are typically present within the cell nucleus. In the future, we plan to examine chromosome territories associations for larger number of pairs and further study the structural differences among different chromosome territory subgroups using the techniques described in this contribution. We believe that the idea of maximum association graph will find general applicability in learning patterns from graph representations, in particular constructing comprehensive three-dimensional arrangement of all 23 pairs of chromosome in a diploid somatic cell nucleus.

References

1. Berezney, R.: Regulating the mammalian genome: the role of nuclear architecture. Advances in Enzyme Regulation 42, 39–52 (2002)
2. Cremer, T., Cremer, C.: Chromosome territories, nuclear architecture and gene regulation in mamalian cells. Nature Reviews Genetics 2, 292–301 (2001)
3. Stein, G.S., Zaidi, S.K., Braastad, C.D., Montecino, M., van Wijnen, A.J., Choi, J.Y., Stein, J.L., Lian, J.B., Javed, A.: Functional architecture of the nucleus: Organizing the regulatory machinery for gene expression, replication and repair. Trends Cell Biol. 13, 584–592 (2003)
4. Jiang, X., Münger, A., Bunke, H.: On median graphs: Properties, algorithms, and applications. IEEE Trans. Pattern Anal. Mach. Intell. 23(10), 1144–1151 (2001)
5. Hlaoui, A., Wang, S.: Median graph computation for graph clustering. Soft. Comput. 10(1), 47–53 (2006)
6. Mukherjee, L., Singh, V., Peng, J., Xu, J., Zeitz, M.J., Berezney, R.: Generalized median graphs: Theory and applications. In: ICCV, pp. 1–8 (2007)
7. Ullmann, J.R.: An algorithm for subgraph isomorphism. J. ACM 23(1), 31–42 (1976)
8. Foggia, P., Sansone, C., Vento, M.: A database of graphs for isomorphism and subgraph isomorphism benchmarking. In: CoRR, pp. 176–187 (2001)
9. Zeitz, M.J., Mukherjee, L., Bhattacharya, S., Xu, J., Berezney, R.: A probabilistic model for the arrangement of a subset of human chromosome territories in wi38 human fibroblasts. Journal of Cellular Physiology 221(1), 120–129 (2009)

Estimation of 3D Geometry of Microtubules Using Multi-angle Total Internal Reflection Fluorescence Microscopy*

Qian Yang[1], Alexander Karpikov[2], Derek Toomre[3], and James Duncan[1,2]

[1] Department of Electrical Engineering
[2] Department of Diagnostic Radiology
[3] Department of Cell Biology
Yale University, New Haven, CT, USA

Abstract. With the ultimate goal to quantify important biological parameters of microtubules, we present a method to estimate the 3D positions of microtubules from multi-angle TIRF data based on the calibrated decay profiles for each angle. Total Internal Reflection Fluorescence (TIRF) Microscopy images are actually projections of 3D volumes and hence cannot alone produce an accurate localization of structures in the z-dimension, however, they provide greatly improved axial resolution for biological samples. Multiple angle-TIRF microscopy allows controlled variation of the incident angle of the illuminating laser beam, thus generating a set of images of different penetration depths with the potential to estimate the 3D volume of the sample. Our approach incorporates prior information about intensity and geometric smoothness. We validate our method using computer simulated phantom data and test its robustness to noise. We apply our method to TIRF images of microtubules in PTK_2 cells and compare the distribution of the microtubule curvatures with electron microscopy (EM) images.

1 Introduction

The study of biological processes of microtubules has been greatly aided by total internal reflection fluorescence (TIRF) microscopy. A microtubule is a hollow cylinder, constructed by 13 protofilaments, about 25 nm in diameter. They have biomedical importance due to their functions as conveyer belts for vesicles inside the cell and to regulate cell migration and division.

The objective-type TIRF microscope has the advantage of commercially accessible and is becoming a powerful tool to study sub-cellular structures. However, we have to calibrate the microscope and measure the axial decay profile. Once properly calibrated, TIRFM provides a high signal to noise ratio by illuminating a limited specimen region immediately adjacent to the substrate. By varying the incident angle, we can obtain TIRF projections with different penetration depths d.

* This work is supported in part by a funding from the KECK Foundation.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 538–545, 2010.

There is some previous work on estimation of depth/distance information from multi-angle TIRF microscopy images. Truskey *et al.* collected TIRF images at different penetration depths d for well-spread bovine aortic endothelial (BAEC) stained with a membrane-bound carbocyanine dye, and determined the depth z by using a simplified model of TIRFM optics and fitting the image intensity versus penetration depths[1]. Ölveczky *et al.* used the inverse Laplace transform to estimate the fluorophore distribution C(x,y,z) and implemented it for a simple geometry and top-hat distribution of fluorohpore(either 0 or 1) to determine cell-substrate distances for the whole x-y plane[2]. More complex structures like spheres were modeled by Rohrbach to estimate the diameters of isolated granules and the distance between vesicles in the solution and the coverslip[3]. Stock *et al.* constructed a four-layer model (glass substrate, aqueous extracellular medium, plasma membrane and the cytoplasm) with different refractive indices and used nonlinear regression to calculate the distance between the plasma membrane and the glass substrate[4]. Estimations of more complicated 3-D structures have not been attempted to date.

In this study we present a novel method that employs multi-angle TIRF microscopy technique to estimate the geometry of 3D curvilinear structures (e.g. microtubules). Fig.1 is an example of multi-angle TIRF images of microtubules with the penetration depths increasing from left to right. Instrumentation was constructed to calibrate the axial profile of the TIRF microscope thus taking into consideration factors that make the profile deviates from theoretically predicted exponential. A maximum a posteriori (MAP) framework is developed to take TIRF images of different penetration depths as input and estimate z positions along segmented microtubules. The method is validated using a computer generated phantom with known geometrical structure and fluorophore distributions. TIRF imaging process is simulated with the phantom and the resulting multi-angle projection images are used as input for 3D estimation. We then apply the method to in vitro microtubules and estimate multiple regions from several data sets.We evaluate the accuracy of the algorithm by imaging microtubules in vitro using different sets of penetration depths and compare the results. The distribution of curvatures along the microtubules are computed and compared with electron microscopy images which are used as ground truth.

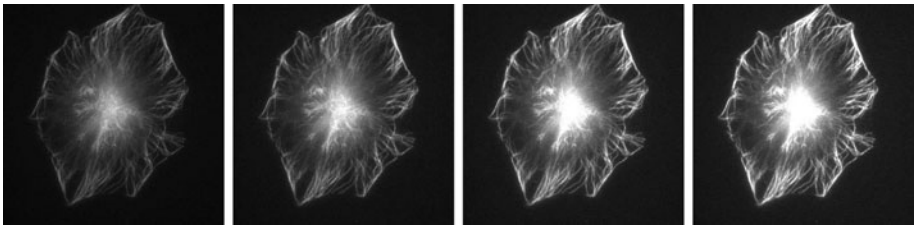


Fig. 1. Multi-angle TIRF images. Left: thinner penetration depth. Right: deeper penetration depth.

2 Methods

2.1 TIRF Imaging Theory

An evanescent field is produced when total internal reflection of a laser beam occurs at the interface between a substrate and a sample[5]. Theoretically, the axial profile of the evanescent field behind the interface follows an exponential decay with a single parameter d, which refers to the penetration depth:

$$I(z, d(\theta)) = I_0(d(\theta)) \cdot e^{-z/d(\theta)} \tag{1}$$

The penetration depth d is associated with the incident angle:

$$d(\theta) = \lambda/4\pi \cdot (n_i^2 \cdot \sin^2\theta - n_t^2)^{-1/2} \tag{2}$$

where n_i and n_t are the refractive indices of the substrate and the sample, θ is the incident angle of the laser beam.

The following formula describes how a TIRF image is formed [3]:

$$I(d(\theta)) = \phi I_0(d(\theta)) \int_0^\infty [Q(z)\mathrm{PSF}(z)]\, C(z) e^{-z/d} \mathrm{d}z \tag{3}$$

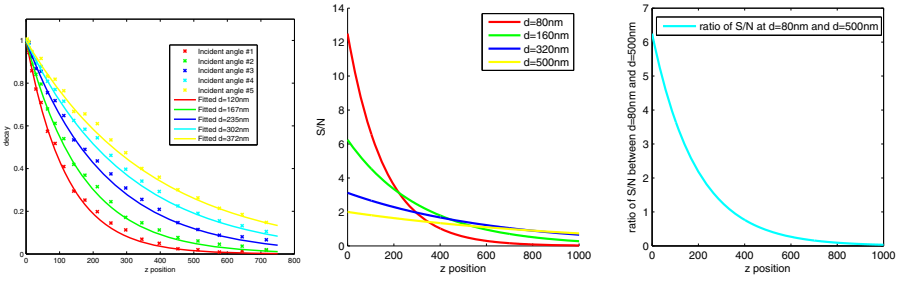
Here I represents the multi-angle TIRF images we record, C is the actual fluorophore concentration, and $I_0(\theta)$ is the intensity directly behind the glass interface. ϕ denotes the quantum efficiency of the fluorophores and CCD camera, $Q(z)$ and $PSF(z)$ are the collection efficiency and point spread function.

To calibrate the axial profile and calculate penetration depth, we place a large spherical silicon bead with known diameter on the glass cover slip and place it in the center of the field of laser illumination[6]. For each specific penetration depth, we can combine the projection image with the geometry of the bead to infer the axial decay profile. The calibration result is shown in Fig.2a. Note that the profiles measured are normalized so we also measured $I_0(d(\theta))$ to compensate.

2.2 Challenges in Multi-angle TIRFM

Assuming a completely constant light source, the photon distribution over a set of bins of equal time is a Poisson probability distribution. Thus, the standard deviation of the number of photon N collected by the CCD camera is $\sqrt{N}$, therefore, the relative uncertainty is given by $\sqrt{N}/N = 1/\sqrt{N}$. This is the major factor leading to falling signal-to-noise ratio along axial direction as we increase the penetration depth to image a thicker part of the sample. Another important factor comes from the background fluorophore in the solution which will also be collected by the camera. Assuming an exponential decay of the axial profile, we calculate the signal-to-noise ratio as a function of depth to be:

$$S/N \propto \left[\frac{N}{\sqrt{N}}\right] \cdot \left[\frac{1}{\int_0^\infty e^{-z/d}\mathrm{d}z}\right] = \sqrt{N} \cdot \frac{1}{d} \propto \frac{e^{-z/2d}}{d} \tag{4}$$



(a) **(b)**

Fig. 2. (a) The decay profiles for different incident angles are calibrated and fitted with exponential decay to find the penetration depths (b) Left: Shot noise/background fluorescence lead to decreasing S/N, with the decreasing rate inversely proportional to the penetration depth. Right: Ratio of S/N between d=80nm and d=500nm shows that the latter has a larger value when z is deeper than 350nm.

where the two terms in the right hand side brackets correspond to the shot noise(photon collecting process) and background fluorescence respectively. This characterizes a major difference between multi-angle TIRF depth estimation from other estimation problems like tomography, where all images are projections of the entire volume and signal-to-noise ratio is the same. With multi-angle TIRFM, there will be a trade-off between the amount of information and the S/N ratio, i.e. as you probe deeper into the sample with increasing penetration depth, S/N ratio will drop rapidly. So while it will be a good idea to use images with deep penetration depth as reference images for segmentation purposes, it will be necessary to revisit structures that also exist in low penetration depth images due to better S/N ratio. While the low penetration depth images provides good S/N ratio for near-the-interface structures, the S/N ratio drops considerably faster than deep penetration depth images (Fig.2), so for very deep structures, they won't provide S/N ratio as good as the latter since the signal intensity deteriorates dramatically.

2.3 MAP Framework with Non-informative Prior

To estimate the z positions for points along the microtubules, we first segment them by manually selecting microtubule tips and finding the shortest path to cells or microtubule organizing centers[7]. Since the point spread function will smear out the microtubule to have a width of several pixels, we fit a Gaussian curve along the perpendicular direction to the microtubule to achieve subpixel accuracy of the actual coordinates[8]. We represent the microtubule centerline represented as $[(x(s), y(s)]$, where s is the curve length. Let M be the number of points sampled along the microtubule, and N be the number of different angles we use to obtain TIRF images. Points where microtubules cross each other are excluded to avoid confusion. A MAP framework is employed to estimate the actual depth information $[z_1, z_2, ..., z_M]$ along the microtubule $[(x_1, y_1), (x_2, y_2), ...(x_M, y_M)]$

from a set of TIRF images $[I_1, I_2, ..., I_N]$ corresponding to different penetration depths with axial decay profiles $[f_1, f_2, ..., f_n]$.

Our objective function is the posterior distribution of the depth z:

$$\begin{aligned} E &= \log p(z|I_1, ..., I_n; f_1, ..., f_N) \\ &\propto \log p(I_1, ..., I_N|z; f_1, ..., f_N)p(z) \\ &= \log \left[\prod_{i=1}^{M} \left(\prod_{j=1}^{N} \left(p(I_j(x_i, y_i)|z_i; f_j) \right)^{w_j(z_i)v(z_i)} \right) p(z_i) \right] \\ &= \log \left[\prod_{i=1}^{M} \left(\prod_{j=1}^{N} \left(\frac{f_j(z_i)^{I_{j,i}} e^{-f_j(z_i)}}{I_{j,i}!} \right)^{w_j(z_i)v(z_i)} \right) p_d(z_i)p_g(z_i) \right] \end{aligned} \tag{5}$$

Breaking up the joint density function $p(I_1, ..., I_N|z; f_1, ..., f_N)$ into products of $p(I_j(z_i)|f_j, z_i)$ is based on an assumption that assumes independence between projection images I_j corresponding to different penetration depths, which arises from independence between f's and z's. $p(I_j(x_i, y_i)|z_i; f_j)$ is the likelihood function that illustrates the model of the CCD camera photo collection process, which is well approximated by a Poisson distribution. The prior distribution is consist of the depth prior p_d and the geometric prior p_g.

The weight coefficient w_j is chosen to be proportional to the S/N ratio corresponding to penetration depth d_j, thus penalizing projections with low reliability and put more weight on

$$w_j(z_i) \propto (S/N)_j(z_i) = \frac{N \cdot e^{-z_i/2d_j}/d_j}{\sum_{j=1}^{N} e^{-z_i/2d_j}/d_j} \tag{6}$$

The weight coefficient v is chosen to be a increasing function of z to compensate the diminishing likelihood functions as z increases. v is determined heuristically to be a piecewise constant function which exhibits more numerical stability.

Since we have no prior information nor preference on the parameter z we wish to estimate, we use non-informative prior (Jeffreys prior) distribution in our model. It has the property to be invariant under reparameterization f_j's. Jeffreys prior is proportional to the square root of the determinant of the Fisher information:

$$p_d(z_i) \propto \sqrt{\det \mathcal{I}(z_i)} = \sqrt{\sum_{j=1}^{N} \frac{(f_j'(z_i))^2}{f_j(z_i)}} \tag{7}$$

It turns out that p_d is a decreasing function of z, which agrees with our intention to encourage near-substrate positions in many biological circumstances (*e. g.* a cell with a leading edge).

Geometric prior is used to add local smoothness constraint along the microtubules [9], α and β control the "elasticity" and "rigidity" of the microtubules:

$$p_g(z_i) = \alpha \left\| \frac{\partial z}{\partial s} \right\|^2_{(x_i, y_i, z_i)} + \beta \left\| \frac{\partial^2 z}{\partial s^2} \right\|^2_{(x_i, y_i, z_i)} \tag{8}$$

Maximizing the objective function results in our estimation of the depth z.

3 Results

3.1 Computer Simulation Using Phantom Data

To validate our estimation method, we generate a $128 \times 128 \times 40$ computer simulated phantom with each voxel of size $160nm \times 160nm \times 50nm$, which corresponds to the resolution of the TIRF microscope. We put curvilinear structures in the volume and simulate the TIRF imaging process by projecting the volume into TIRF projection images using different penetrations depths. Background noise is modeled as the ratio of standard deviation of fluorophore concentration in solution and the concentration on the object of interest (assumed to be uniform). Poisson distribution is used to model the photon collecting process of the CCD camera which generates a TIRF image.

We use the set of projection images as input to our method and estimate the depth of the tubular objects. The optimal number of angles used is a tradeoff between single frame estimation accuracy and time resolution of image sequences. Assuming a constant background noise level, we estimate the volume using different number of input angles. As we can see in Fig.3, estimation accuracy using only one input angle is unacceptable. Basically, our algorithm just turn into a trivial case of solving an equation like $I = I_0 exp(-z/d)$ if there is just one projection image for a single incident angle. When the number of angle is equal to 2, we begin to see improvements for shallow structures near the interface. Difference between the results from 5 and 10 input angles is not visually detectable.

Robustness is studied by computing depth estimation error versus noise level and depth estimation error versus depth of object of interest. Resulting plots show a linear relation between the estimation error and background noise, which suggests that our algorithm is quite robust to noise, especially when the structures being studied is close to the glass interface. While the target structures get deeper into the evanescent field, we would expect a linear deterioration of estimation accuracy. Fortunately, most of the biological structures we study(plus end of the microtubules) are close to the substrate and generally won't exceed a maximum depth of $1000nm$.

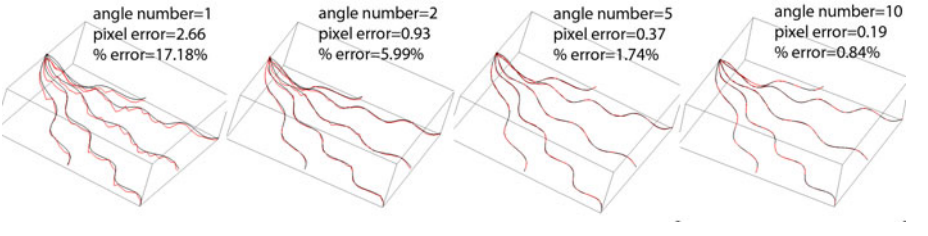


Fig. 3. Estimation error vs number of input angles. Black: Original phantom. Red: Estimated structures.

3.2 Estimation Results for in vitro Microtubules

In our experiments TIRFM images were acquired using an inverted microscope equipped with a high numerical aperture (NA=1.495, 60X) oil-immersion lens (Olympus) and a back-illuminated electron-multiplying charge-coupled device camera (512 x 512, 16-bit; iXon887; Andor Technologies), and controlled by Andor iQ software (Andor Technology). Excitation was achieved using a 514nm line of argon laser, and exposure times were 0.1-0.2 s. PTK_2 cells were fixed with 2.5% paraformaldehyde and 3% glutaraldehyde in 0.1 M cacodylate buffer (pH 7.4) for 2 hours at room temperature, then postfixed for 2 hours at room temperature with 1% OsO_4 in 0.1 M cacodylate buffer (pH 7.4), dehydrated through a series of increasing ethanol concentrations.

Fig.4a and Fig.4d present the estimation results for 5 individual microtubules in a peripheral area of a cell. The z axis in Fig.4d is exaggerated for demonstration purpose. In Fig.4b we crop out multiple regions of a cell, and estimate them at the same time. Fig.4e shows the 3D view of the estimated volume.

To further validate our method, we pick a single microtubule and repeat the estimation process using different sets of penetration depths. Fig.4c shows that the estimation results differ by less than the diameter of the microtubules (25nm), which is considered highly precise. Since the penetration depths and

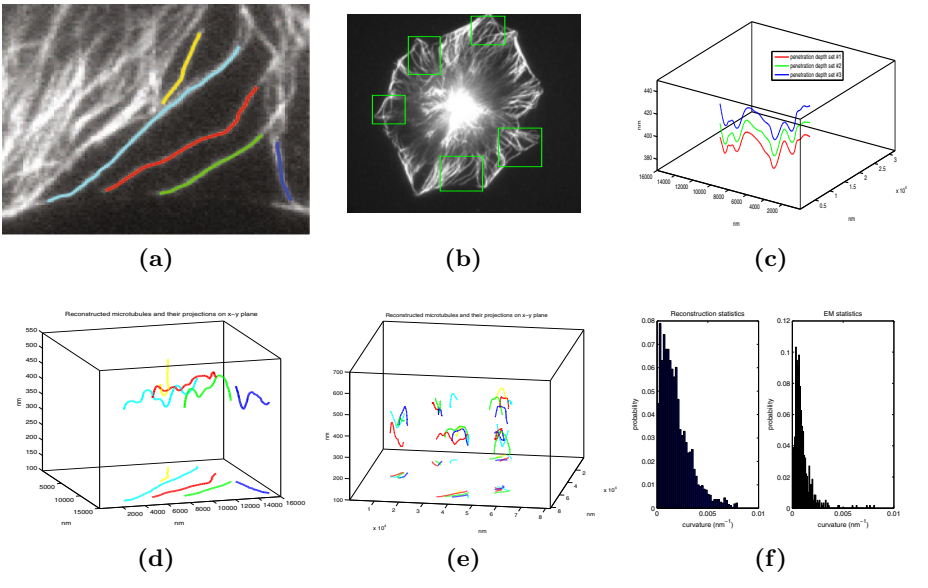


Fig. 4. (a)&(d) Estimation results from a single peripheral region. (b)&(e) Estimation results from multiple peripheral region. A magnified view of the bottom region is shown in (a). (c) Comparison of estimation results using different sets of penetration depth and TIRF images. (f) Comparison of histograms of curvatures from the estimation (left) and EM data (right).

their corresponding TIRF images are acquired independently, the fact that the estimation results agree with each other further validate our algorithm. Fig.4f compares the distribution of curvatures along the estimated microtubules and in the electron microscopy images we use as a ground truth due to its much larger magnification and the ability to image an axial section. We can see they follow similar distributions, and mostly important, have the same maximum curvature, which is a significant characteristic of microtubules.

4 Discussion

The purpose of this study is to develop and validate a method based on MAP framework and optimization techniques to estimate 3-D structures from a set of 2-D projection images generated using multi-angle TIRF microscopy. Our method takes the measured decay curves and corresponding projection images as input, and estimates the axial position of the objects of interest. Computer simulations are used to validate the noise robustness of the algorithm, and also determine the optimal number of input angles we should use. We apply the method to in vitro microtubule images and obtain satisfactory results. The estimation method presented here will be useful for quantifying useful parameters related to microtubules' behavior (movement, growth, etc).

References

1. Truskey, G.A., Burmeister, J.S., Grapa, E., Reichert, W.M.: Total internal reflection fluorescence microscopy (tirfm) ii. topographic mapping of relative cell/substrate separation distances. Journal of Cell Science 103, 491–499 (1992)
2. Ölveczky, B.P., Periasamy, N., Verkman, A.S.: Mapping fluorophore distributions in three dimensions by quantitative multiple angle-total internal reflection fluorescence microscopy. Biophysics Journal 73, 2836–2847 (1997)
3. Rohrbach, A.: Observing secretory granules with a multiangle evanescent wave microscopy. Biophysical Journal 78, 2641–2654 (2000)
4. Stock, K., Sailer, R., Strauss, W.S.L., Lyttek, M., Steiner, R., Schneckenburger, H.: Variable-angle total internal reflection fluorescence microscopy(va-tirfm): realization and application of a compact illumination device. Journal of Microscopy 211, 19–29 (2003)
5. Axelrod, D., Hellen, E.H., Fulbright, R.M.: Total internal reflection fluorescence. Topics in fluorescence spectroscopy: Principles and applications 3, 289–343 (1992)
6. Mattheyses, A.L., Axelrod, D.: Direct measurement of the evanescent field profile produced by objective-based total internal reflection fluorescence. Journal of Biomedical Optics 11(1) (2006)
7. Cohen, L.D., Kimmel, R.: Global minium for active contour models: A minimum path approach. International Journal of Computer Vision 24(1), 57–78 (1997)
8. Bicek, A.D., Tüzel, E., Demtchouk, A., Uppalapati, M., Hancock, W.O., Kroll, D.M., Odde, D.J.: Anterograde microtubule transport drives microtubule bending in llc-pk1 epithelial cells. Molecular Biology of the Cell 20, 2943–2953 (2009)
9. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. In: International Journal of Computer Vision, 321–331 (1988)

Recursive Green's Function Registration

Björn Beuthien[1,2,3], Ali Kamen[4], and Bernd Fischer[1,2]

[1] Institute of Mathematics and Image Computing, University of Lübeck, Germany
[2] Fraunhofer MEVIS, Project Group Image Registration, Lübeck, Germany
[3] Graduate School for Computing in Medicine and Life Sciences, Lübeck, Germany
[4] Siemens Corporate Research, Princeton NJ, USA

Abstract. Non-parametric image registration is still among the most challenging problems in both computer vision and medical imaging. Here, one tries to minimize a joint functional that is comprised of a similarity measure and a regularizer in order to obtain a reasonable displacement field that transforms one image to the other. A common way to solve this problem is to formulate a necessary condition for an optimizer, which in turn leads to a system of partial differential equations (PDEs). In general, the most time consuming part of the registration task is to find a numerical solution for such a system. In this paper, we present a generalized and efficient numerical scheme for solving such PDEs simply by applying 1-dimensional recursive filtering to the right hand side of the system based on the Green's function of the differential operator that corresponds to the chosen regularizer. So in the end we come up with a general linear algorithm. We present the associated Green's function for the diffusive and curvature regularizers and show how one may efficiently implement the whole process by using recursive filter approximation. Finally, we demonstrate the capability of the proposed method on realistic examples.

Keywords: Nonparametric Image Registration, Green's Function, Recursive Filter.

1 Introduction

The problem of image registration arises in many application of medical image processing and computer vision. Given two images, a reference R and a template T, we try to find a suitable transformation that aligns the template to the reference.

In this paper, we focus on non-parametric image registration, where we minimize a joint functional depending on a (dis)similarity measure and a regularizer. The optimization leads to a system of partial differential equations, the so-called Euler-Lagrange equations. In the literature one may find various ways for solving the PDEs, see [7] for an overview. In this paper, we introduce a generalization of solving this problem by convolution. For speed purposes, this convolution is approximated by 1-dimensional recursive filtering. The paper is organized as follows. We start by describing the registration problem in more detail followed by a

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 546–553, 2010.

short introduction to the Green's function approach. In particular we comment on an efficient way to determine the Green's function based on eigenfunction expansion. In the following we come up with a recursive filter approximation approach for the presented filters and present the resulting algorithm. Finally, we provide numerical examples for CT lung data.

2 Image Registration

In this section we will state the mathematical formulation for the non-parametric image registration problem and give an overview of well-known methods. For d-dimensional images $R, T : \mathbb{R}^d \to \mathbb{R}$, we are looking for a non-parametric transformation $\phi(\mathbf{x}) = \mathbf{x} - \mathbf{u}(\mathbf{x})$, where $\mathbf{u} : \mathbb{R}^d \to \mathbb{R}$ is such that the deformed template $T_u(\mathbf{x}) := T(\mathbf{x} - \mathbf{u}(\mathbf{x}))$ is similar to the reference R. In this context, the vector field $\mathbf{u}$ is called displacement field. This problem can be solved by minimizing the joint functional

$$\mathcal{J}[\mathbf{u}] = \mathcal{D}[R, T; \mathbf{u}] + \alpha \mathcal{S}[\mathbf{u}], \tag{1}$$

where $\mathcal{D}$ is a image similarity or distance measure and $\mathcal{S}$ is a regularization or smoothing term. The smoothing term constrains the transformation to a set of "reasonable" ones, which may be used to advantage for the given application. The regularization parameter α controls the relative contributions of the two terms.

2.1 Distance Measures

To measure similarity of the two images we have to define an appropriate distance measure $\mathcal{D}$. Frequently used measures are the sum of squared differences (SSD), normalized gradient fields [5] or mutual information [12]. In this paper, we will exemplarily use SSD, that is given by

$$\mathcal{D}^{\mathrm{SSD}}[R, T; \mathbf{u}] := \int_\Omega \left(T_u(\mathbf{x}) - R(\mathbf{x})\right)^2 \, \mathrm{d}x. \tag{2}$$

2.2 Regularizers

Basically regularizers measure the smoothness of the wanted the transformation. Popular regularization terms are defined as follows:

$$\mathcal{S}^{\mathrm{diff}}[\mathbf{u}] := \frac{1}{2} \sum_{j=1}^{d} \int_\Omega \|\nabla u_j(\mathbf{x})\|^2 \, d\mathbf{x}, \tag{3}$$

$$\mathcal{S}^{\mathrm{curv}}[\mathbf{u}] := \frac{1}{2} \sum_{i=1}^{d} \int_\Omega \left(\Delta u_l(\mathbf{x})\right)^2 \, d\mathbf{x}, \tag{4}$$

$$\mathcal{S}^{\mathrm{elas}}[\mathbf{u}] := \int_\Omega \frac{\mu}{4} \sum_{j,k=1}^{d} \left(\partial_{x_j} u_k + \partial_{x_k} u_j\right)^2 + \frac{\lambda}{2} \left(\nabla \cdot \mathbf{u}\right)^2 \, d\mathbf{x}. \tag{5}$$

For fluid registration $\mathcal{S}^{\text{fluid}}[\mathbf{u}] := \mathcal{S}^{\text{elas}}[\mathbf{v}]$. In this context ∂_{x_j} denotes the partial derivative in direction x_j, ∇ the gradient, $\nabla\cdot$ the divergence and Δ the Laplacian operator. Furthermore $\mathbf{v} : \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the velocity field and λ, μ are called Lamé constants. The regularization terms shown above are known as diffusive, curvature, elastic, and fluid registration, respectively (see [7]).

2.3 Numerical Solutions

For an optimal displacement field $\mathbf{u}$ the Gâteaux derivatives of the joint functional in Equation (1) vanishes. This leads to the corresponding set of non-linear Euler-Lagrange equations

$$\alpha\mathcal{A}[\mathbf{u}] = \mathbf{f}(\mathbf{x}, \mathbf{u}(\mathbf{x})), \tag{6}$$

where the differential operator $\mathcal{A}$ and the force $\mathbf{f}$ are the Gâteaux derivatives of $\mathcal{S}$ and $\mathcal{D}$, respectively (under the assumption of specific boundary conditions). The resulting differential operators for the various regularization terms (3), (4), and (5) are

$$\mathcal{A}^{\text{diff}} = \Delta, \tag{7}$$

$$\mathcal{A}^{\text{curv}} = \Delta^2 \text{ and} \tag{8}$$

$$\mathcal{A}^{\text{elas}} = \mu\Delta + (\lambda + \mu)\,\nabla\nabla\cdot. \tag{9}$$

The differential operator for fluid registration is similar to the elastic one, but operates on the velocity field as opposed to the displacement field as discussed in e.g. [7]. A fixed-point-type iteration scheme, such as

$$\mathcal{A}[\mathbf{u}^{k+1}] = \mathbf{f}(\mathbf{x}, \mathbf{u}^k(\mathbf{x})), \tag{10}$$

is a practical way to linearize and solve these equations. We have various options in providing a numerical scheme for first discretizing and solving the PDEs. The main work is the solution of the resulting linear system. This can be done by successive overrelaxation (SOR) [3], multigrid methods [6], additive operator splitting (AOS) for diffusive regularization and homogeneous Neumann boundary conditions [7] or by Fourier methods [2], to name few popular options. An alternative method, as suggested by Bro-Nielsen and Gramkow, for fluid registration [1] is to solve the Euler-Lagrange equations by means of the convolution operation

$$\mathbf{u}^{k+1} = G * \mathbf{f}^k, \tag{11}$$

where G is an appropriate filter kernel which turns out to be the Green's function w.r.t. the differential operator under the assumption of translational invariance. In the following we will assume this condition is fulfilled although boundary conditions are applied. This approach can be generalized to other regularizers as well.

3 Green's Function

The Green's function is used to solve linear inhomogeneous ordinary or partial differential equations, e.g. in electromagnetic theory [10]. We consider sufficiently smooth functions $\mathbf{u}, \mathbf{f}, \mathbf{g} : \mathbb{R}^d \to \mathbb{R}^d$, an open set $\Omega \subset \mathbb{R}^d$ with smooth boundary $\partial\Omega$ and the linear operator $\mathcal{A}$ for the Euler-Lagrange equations (6). Therefore, we have

$$\mathcal{A}\mathbf{G}^i(\mathbf{x};\mathbf{y}) = \delta(\mathbf{x}-\mathbf{y})\mathbf{e}_i \qquad \text{for } \mathbf{x} \in \Omega \tag{12}$$

$$\mathbf{G}^i(\mathbf{x};\mathbf{y}) = \mathbf{g}(\mathbf{x}) \qquad \text{for } \mathbf{x} \in \partial\Omega, \tag{13}$$

where $\mathbf{G}^i = \left(G_1^i \dots G_d^i\right)^T$, is called *vector Green's function* in the direction $\mathbf{e}_i, i \in \{1,\dots,d\}$. Then $\mathbf{G}(\mathbf{x};\mathbf{y}) := \left(\mathbf{G}^1(\mathbf{x};\mathbf{y}) \dots \mathbf{G}^d(\mathbf{x};\mathbf{y})\right)$ is called *dyadic Green's function*, using which the following equation must hold:

$$\mathcal{A}[\mathbf{G}](\mathbf{x};\mathbf{y}) = \mathbf{I} \cdot \delta(\mathbf{x}-\mathbf{y}), \tag{14}$$

where $\mathbf{I} \in \mathbb{R}^{d\times d}$ is the identity.

3.1 Green's Function Calculation

To the best of our knowledge, there is no established and commonly accepted way to derive Green's function for a differential operator. Here, we will introduce the method of eigenfunction expansion.

Eigenfunction Expansion. The idea of the eigenfunction expansion is to express the solution of a differential equation by a weighted sum of orthonormal eigenfunctions Φ_i, i.e.

$$u(\mathbf{x}) = \sum_{i=1}^{\infty} a_i \Phi_i(\mathbf{x}), \quad a_i = \frac{\Phi_i(\mathbf{y})}{\kappa_i}, \tag{15}$$

where κ_i is the eigenvalue that corresponds to the eigenfunction Φ_i (see e.g. [1]). Equation (15) leaves us with two problems. How to compute the eigenfunctions and eigenvalues for a given operator and how to deal with the infinite sum. For the latter problem an answer is also given in [1]. In the following, we present the eigenfunction expansions for diffusive and curvature registration for the 2-dimensional case. The extension to the 3-dimensional case is straightforward. The Green's function for the elastic regularizer coincide with the one of the fluid registration. The only difference is the fact that it is applied to the displacement instead of the velocity.

Diffusive Registration. The Green's function for the diffusive operator $\mathcal{A}^{\text{diff}}$ under zero boundary conditions on the domain $\Omega =]0,1[^2$ is given by a scalar Green's function

$$G(\mathbf{x};\mathbf{y}) = -\sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \frac{\Phi(\mathbf{y})}{\kappa^{k,l}} \Phi(\mathbf{x}). \tag{16}$$

with eigenfunctions and eigenvalues

$$\Phi^{k,l}(\mathbf{x}) = \sin(k\pi x_1)\sin(l\pi x_2) \quad \text{resp.} \quad \kappa^{k,l} = -\pi^2\left(k^2+l^2\right). \tag{17}$$

One can shown that the diffusive operator is separable and therefore the displacement field can be computed by $u_1(\mathbf{x}) = (G * f_1)(\mathbf{x})$ and $u_2(\mathbf{x}) = (G * f_2)(\mathbf{x})$.

Curvature Registration. As in the case of the diffusive registration, the operator for the curvature registration is separable as well. We also consider zero boundary conditions and the domain $\Omega =]0,1[^2$. The eigenfunctions and eigenvalues are given by

$$\Phi^{k,l}(\mathbf{x}) = \sin(k\pi x_1)\sin(l\pi x_2) \quad \text{resp.} \quad \kappa^{k,l} = \pi^4\left(k^2+l^2\right)^2. \tag{18}$$

3.2 Deriving a Recursive Filter

In this work the Green's function $G(\mathbf{x};\mathbf{y})$ is given in its continuous eigenfunction representation. The force field is given by Equation 2. Both of them are discretized on a finite grid. We get the Green's filter $\mathbf{G}$ and the discrete force field $\mathbf{F}$. The discrete displacement field $\mathbf{U}$ is now given by the convolution $\mathbf{U} = \mathbf{G} * \mathbf{F}$. This operation has high computational costs, i.e. $O(N^2)$, where N is the number of voxel. To reduce these costs the idea of 1-dimensional recursive filtering will be introduced. In the following we will compute a separable filter approximation where the recursive filtering in each direction can be handled step by step. This allows us to use an implementation of diffusive and curvature registration with the computational costs of the demons approach using a recursive implementation of the Gaussian.

We start with the 2-dimensional case. In terms of computing a separable recursive filter approximation for the Green's filter $\mathbf{G} \in \mathbb{R}^{n_1 \times n_2}$, we need a separable approximation $\widetilde{\mathbf{G}} \in \mathbb{R}^{n_1 \times n_2}$ of $\mathbf{G}$ of the form

$$\widetilde{\mathbf{G}} = \mathbf{x} \otimes \mathbf{y}, \quad \text{resp.} \quad \widetilde{\mathbf{G}} = \mathbf{x}\mathbf{y}^T, \tag{19}$$

where $\mathbf{x} \in \mathbb{R}^{n_1}$ and $\mathbf{y} \in \mathbb{R}^{n_2}$. This is the case if $\operatorname{rank}(\widetilde{\mathbf{G}}) = 1$. We choose this approximation to be optimal in terms of the Frobenius norm ($\|\cdot\|_F$) that computations are easy to handle. This optimal approximation is called a rank-one approximation and can be achieved by a singular value decomposition (SVD) [4]. A singular value decomposition is possible for all matrices $\mathbf{G}$. Let $\mathbf{G} = \mathbf{U}\mathbf{S}\mathbf{V}^T$, where $\mathbf{U} = \left(\mathbf{u_1} \dots \mathbf{u_{n_1}}\right) \in \mathbb{R}^{n_1 \times n_1}$ and $\mathbf{V} = \left(\mathbf{v_1} \dots \mathbf{v_{n_2}}\right) \in \mathbb{R}^{n_2 \times n_2}$ are orthogonal matrices and $\mathbf{S} \in \mathbb{R}^{n_1 \times n_2}$ is a diagonal matrix. Then $\widetilde{\mathbf{G}} = \mathbf{u}\mathbf{v}^T = \left(\sqrt{s_{11}}\mathbf{u}_1)\right)\left(\sqrt{s_{11}}\mathbf{v}_1)\right)^T$. For the 3-dimensional case the application of the SVD is not possible in this way. A rank-one approximation of a third order tensor can be seen as a generalization of the SVD and is often called multidimensional SVD. This problem can be attacked by the generalized Rayleigh-Newton iteration as described by Zhang and Golub [13]. The separable approximation $\widetilde{\mathbf{G}}$ is then given by $\widetilde{\mathbf{G}} = \mathbf{x} \otimes \mathbf{y} \otimes \mathbf{z}$.

The d-dimensional convolution can now be substituted by d 1-dimensional convolutions what is already a nice speed-up. But the 1-dimensional convolution has still high costs. That is where recursive filtering becomes an issue. The recursive filter approximation of the 1-dimensional filters is done by using Matlab implementation of Prony's method [8] with a zero padding to emphasize the zero-boundary conditions in the filter. It should be mentioned that the combination of causal and anti-causal recursive filter approximation does not lead to a perfect symmetry in this implementation. However, in our case this aspect is hardly noticeable. This recursive filtering scheme can now be implemented with O(N) computation cost.

4 Numerical Examples

In the following evaluate the capability of the Green's function registration on the POPI breathing thorax model of the Léon Bérard Cancer Center & CREATIS lab, Lyon, France [11]. Before starting the evaluation we leave some words on the algorithm. We modified the fixed-point type iteration scheme (Equation (10)). Instead of computing the whole displacement field in one iteration, we compute update steps $\mathbf{u}^k$ by recursive filtering the force field. The final displacement field after n iteration steps is then given by $\mathbf{u} = \sum_{k=1}^{n} \mathbf{u}^k$. The modification is already mentioned in [7] for elastic registration. To have more control over the iteration step we included an Armijo line search for these updates. To allow larger deformations we start with an affine-linear preregistration followed by a multi-level Green's function registration. The choice of the distance measure is the SSD (Equation 2). This is not the best choice for the registration of lung volumes but the aim of this paper is to show the capability of the proposed linear algorithm. Stopping criteria on each level is a difficult topic. Here, we consider the relative distance measure $\bar{d}_k = \mathcal{D}^{\mathrm{SSD}}(R, T; \mathbf{u}^k)/\mathcal{D}^{\mathrm{SSD}}(R, T; \mathbf{0})$. The algorithm stops when the change of the relative distance $\bar{d}$ is less than 1%, i.e. $\bar{d}_k - \bar{d}_{k-1} < 0.01$. Furthermore a maximum number of iterations is given that is descending as the level is ascending.

The POPI-model provides a 4D-CT series of the lung, including ten 3D-CT volumes $(v_0, \ldots, v_9)$ representing ten different phases of an average breathing cycle. For evaluation purposes a set of corresponding landmarks is given for all these volumes. For the registration the preprocessed versions of these volumes with removed background and reduced image size are used. This results in an image size of 482 x 360 x 141 with resolution 0.976562mm $\times$ 0.976562mm $\times$ 2mm. Our registration works on an image size of $256 \times 256 \times 128$ computed by linear interpolation. Therefore the resulting working resolution on the finest level is 1.8387mm $\times$ 1.3733mm $\times$ 2.2031mm. We perform a curvature and a diffusive recursive Green's function registration on these data. An exemplarily result for the diffusive registration is shown in Figure 2. To coincide with the results presented with the model computed with a demon-based approach, we use the second volume v_1 as the reference and all others as templates. For each registration the mean and standard deviation of the 40 landmarks are given for

Reg	Data μ/σ	pTRE μ/σ	cTRE μ/σ	dTRE μ/σ	pMax	cMax	dMax
10	0.48 / 0.54	1.3 / 0.3	0.67 / 0.35	0.68 / 0.34	1.8	1.79	1.75
12	0.49 / 0.62	1.4 / 0.2	0.71 / 0.37	0.69 / 0.42	2.1	1.69	1.57
13	2.19 / 1.82	1.4 / 0.4	1.58 / 0.89	1.33 / 0.70	2.3	3.63	2.70
14	4.33 / 2.51	1.2 / 0.4	1.42 / 0.83	1.20 / 0.73	2.3	5.03	4.32
15	5.75 / 2.64	1.3 / 0.5	1.50 / 0.99	1.42 / 0.92	2.6	6.27	5.48
16	6.10 / 2.92	1.1 / 0.4	1.43 / 0.84	1.30 / 0.72	2.0	5.29	4.20
17	5.03 / 2.33	1.3 / 0.5	1.42 / 0.88	1.34 / 0.76	2.4	4.66	3.56
18	3.68 / 1.57	1.1 / 0.3	1.12 / 0.80	1.12 / 0.72	1.7	3.76	3.71
19	2.07 / 1.06	1.1 / 0.3	0.95 / 0.68	0.88 / 0.58	1.9	3.18	2.97
Total	3.35 / 1.78	1.2 / 0.4	1.20 / 0.74	1.17 / 0.66	2.6	6.27	5.55

Fig. 1. Results of the recursive curvature Green's function registration on the POPI-model in mm. The row with the registration index v_1v_2 presents the registration result of the registration of volume v_1 and volume v_2. Presented are the mean and standard deviation of the initial landmarks (second column) and the TRE of the registration presented with the POPI-model (pTRE, third column), the curvature registration (cTRE, fourth column) and diffusive registration (dTRE, fifth column). The last three columns show the maximal TREs for POPI, curvature and diffusive.

the initial situation as well as the mean and standard deviation of the TRE. Additionally the maximal TRE is presented (see Figure 1).

Considering the working resolution these results are very satisfying. The mean μ of the TRE of the three methods is nearly the same. However, the standard deviation σ recursive Green's function registration is larger. This might have two reasons. Firstly, the errors occur mainly in the x_3-component where we use a lower resolution. The other and probably more important reason results from the used distance measure. As mentioned before we apply the SSD without further preprocessing to take the lung density change into account as proposed in [9]. To achieve better results, a next step would be to incorporate this preprocessing or to use a more suitable distance measure for this specific application. This would also help to catch the outliers. Another important information from the table is the fact that the curvature Green's function registration produces less accurate but smoother results than the diffusive one, so the behavior of the regularizers are preserved in this method.

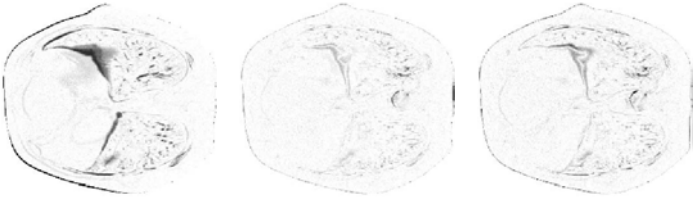


Fig. 2. The absolute difference of the template and the reference before (left) and after the diffusive (middle) and curvature (right) recursive Green's function registration for an exemplarily slice

5 Conclusions

This papers presents a generalization of the registration by a convolution with the Green's function. It was shown that also for the diffusive and curvature regularization filter kernels can be established and recursively approximated. This helps us to speed up the computation in such a way a linear algorithm could be presented. The numerical section presents promising results of this approach by registering a set of CT lung data. Next steps include e.g. the search for the best filter kernel size on each level and the use of application specific distance measures.

References

1. Bro-Nielsen, M., Gramkow, C.: Fast Fluid Registration of Medical Images. In: Höhne, K.H., Kikinis, R. (eds.) VBC 1996. LNCS, vol. 1131, pp. 267–276. Springer, Heidelberg (1996)
2. Cahill, N.D., Noble, J.A., Hawkes, D.J.: Fourier methods for nonparametric image registration. In: CVPR. IEEE Computer Society, Los Alamitos (2007)
3. Christensen, G.E.: Deformable Shape Models for Anatomy. PhD thesis, Washington University (1994)
4. Golub, G.H., Van Loan, C.F.: Matrix Computations. Johns Hopkins Studies in Mathematical Sciences, 3rd edn. The Johns Hopkins University Press, Baltimore (1996)
5. Haber, E., Modersitzki, J.: Intensity Gradient Based Registration and Fusion of Multi-Modal Images. Methods of Information in Medicine 46(3), 292–299 (2007)
6. Hermosillo, G.: Variational Methods for Multimodal Image Matching. PhD thesis, Universite de Nice, France (2002)
7. Modersitzki, J.: Numerical Methods for Image Registration. Oxford Science Publications, Oxford (2004)
8. Parks, T.W., Burrus, C.S.: Digital Filter Design. Wiley, Chichester (1987)
9. Sarrut, D., Boldea, V., Miguet, S., Ginestet, C.: Simulation of 4D CT images from deformable registration between inhale and exhale breath-hold CT scans. Med. Phys. 33(3), 605–617 (2006)
10. Tai, C.-T.: Dyadic Green Functions in Electromagnetic Theory, 2nd edn. Institute of Electrical & Electronics Engineering (1994)
11. Vandemeulebroucke, J., Sarrut, D., Clarysse, P.: Point-validated pixel-based breathing thorax model. In: International Conference on the Use of Computers in Radiation Therapy (ICCR), Toronto, Canada, p. 6247 (June 2007)
12. Viola, P.A.: Alignment by Maximization of Mutual Information. PhD thesis, Massachusetts Institute of Technology (1995)
13. Zhang, T., Golub, G.H.: Rank-one approximation to high order tensors. SIAM Journal on Matrix Analysis and Applications 23(2), 534–550 (2002)

Summarizing and Visualizing Uncertainty in Non-rigid Registration

Petter Risholm[1,2], Steve Pieper[1], Eigil Samset[2], and William M. Wells III[1]

[1] Harvard Medical School, Brigham & Women's Hospital, US
[2] Center of Mathematics for Applications, University of Oslo, NO
pettri@bwh.harvard.edu

Abstract. Registration uncertainty may be important information to convey to a surgeon when surgical decisions are taken based on registered image data. However, conventional non-rigid registration methods only provide the most likely deformation. In this paper we show how to determine the registration uncertainty, as well as the most likely deformation, by using an elastic Bayesian registration framework that generates a dense posterior distribution on deformations. We model both the likelihood and the elastic prior on deformations with Boltzmann distributions and characterize the posterior with a Markov Chain Monte Carlo algorithm. We introduce methods that summarize the high-dimensional uncertainty information and show how these summaries can be visualized in a meaningful way. Based on a clinical neurosurgical dataset, we demonstrate the importance that uncertainty information could have on neurosurgical decision making.

1 Introduction

An important, but somewhat neglected, topic in the field of non-rigid registration is: how can we quantify and visualize the registration uncertainty. The importance can be easily understood in the realm of neurosurgery. Important functional areas of the brain are commonly defined in the pre-operative domain. Whenever these eloquent areas are located adjacent to a tumor, it is critical that their location are accurately mapped intra-operatively. Current image guided navigation systems use rigid registration to establish the mapping between the pre- and intra-operative space, but we are beginning to see a shift towards non-rigid registration which is necessary to accommodate for the non-rigid movement of brain-tissue caused by for instance brain-shift. Conventional registration methods estimate the most likely deformation and consequently the most likely location of a functional area. However, non-rigid registration can behave chaotically in the sense that small changes of input parameters can yield dramatically different results. Some results may be categorized as "registration errors" while others represent viable alternate explanations of the underlying physical reality based on incomplete imaging data. Furthermore, factors such as resection of tissue, degraded intra-operative image quality and blood and edema around a tumor, can all contribute to an increase in the registration uncertainty

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 554–561, 2010.

in the vicinity of tumors. Hence, for a neurosurgeon, the uncertainty in the estimated location of functional areas can be just as important information as the most likely estimate of the location.

Few authors have attempted to quantify the full uncertainty of the estimates produced by non-rigid registration methods. Hub et al. [1] developed a heuristic method for quantifying uncertainty in B-spline image registration by perturbing the B-spline control points and analyzing the effect on the similarity criterion. Kybic [2] consider images to be random processes and proposed a general bootstrapping method to compute statistics on registration parameters. It is a computationally intensive approach, but is applicable to a large set of minimization based registration algorithms. In [3], Kybic assumes normally distributed registration parameters and use the Hessian of the similarity criterion to compute confidence intervals of the registration parameters. Some reported applications of uncertainty estimation in other medical applications include; visualization of uncertainty in image guided needle insertions [4], uncertainty in dose control in radiation oncology [5] and determining uncertainty in fiber orientation of diffusion tensors [6].

In contrast to the frequentist approach of Kybic [2], we presented in [7] an elastic non-rigid Bayesian registration framework that estimates the posterior on both deformation and elastic parameters. We were, firstly, interested in distinguishing different tissue types based on the posterior distribution on elastic parameters, secondly, in analyzing how simultaneous estimation of elastic and registration parameters could potentially improve registration results.

The emphasis of this paper is that, 1) non-rigid registration results come with a level of uncertainty and 2) conveying the uncertainty, and not only the most likely estimate, is clinically important. We characterize the posterior distribution on deformation parameters with the framework in [7] and show several techniques for summarizing and visualizing the uncertainty of the estimated posterior distribution.

2 A Bayesian Framework for Elastic Image Registration

This section provides a short summary of the registration framework in [7]. Let $\mathbf{f}(\mathbf{x})$, $\mathbf{m}(\mathbf{x})$, $\mathbf{x} \in \Omega$ be a fixed (pre-operative) and moving (intra-operative) d-dimensional image respectively with $\Omega \subset \mathbb{R}^d$. We estimate, with a Bayesian model, a deformation $\mathbf{t}(\mathbf{x})$, $\mathbf{x} \in \Omega$, such that $\mathbf{m}(\mathbf{t}(\mathbf{x}) + \mathbf{x})$ is similar to $\mathbf{f}$.

2.1 Bayesian Generative Model

Both the moving image and the deformation are modeled as random variables, while the fixed image is a model parameter. This leads to the following joint probability model, $\mathrm{p}(\mathbf{m}, \mathbf{t}) = \mathrm{p}(\mathbf{m}|\mathbf{t})\mathrm{p}(\mathbf{t})$, where we assume that the observed moving image is dependent on the deformation. According to the theorem of conditional probability, we can write the posterior as:

$$\mathrm{p}(\mathbf{t}|\mathbf{m}) = \frac{\mathrm{p}(\mathbf{m}, \mathbf{t})}{\mathrm{p}(\mathbf{m})} = \frac{\mathrm{p}(\mathbf{m}|\mathbf{t})\mathrm{p}(\mathbf{t})}{\mathrm{p}(\mathbf{m})} \ . \tag{1}$$

We model the likelihood, $p(\mathbf{m}|\mathbf{t})$, with a Boltzmann distribution using a temperature, T_s, and a Sum of Squared Difference (SSD) similarity energy function. Similarly, the prior on the deformation, $p(\mathbf{t})$, is modeled with a Boltzmann distribution with temperature, T_r, and a linear elastic energy function. The elastic energy acts as a regularizer on the deformation and is controlled by the Lamé parameters μ and λ. In [7], we were predominantly interested in generating posterior distributions on the elastic parameters, while this work focuses on the uncertainty of the deformation parameters. Hence, we keep the elastic parameters fixed as model parameters. The energy functions were computed with a linear elastic Finite Element (FE) model.

2.2 Markov Chain Monte Carlo (MCMC) Simulation

Unfortunately, it is not possible to analytically compute the posterior in Eq. (1), nor feasible to draw deformation samples directly from it. A common approach to generate samples from intractable posterior distributions is by way of the Metropolis-Hastings (MH) algorithm [8]. Candidate samples are drawn from a simple proposal distribution, $\mathbf{t}^* \sim N(\mathbf{t}, \sigma_t)$, in our case a Normal distribution, and accepted or rejected according to an acceptance criteria which guarantees that we generate random samples from the posterior distribution. In contrast to [7], where deformations were only sampled for boundary nodes, we sample deformations for all nodes in the FE-mesh. Consequently, we build up a posterior distribution on deformations, as well as marginal posterior distributions on the movements of individual nodes.

The sampling can easily be extended to reject samples that lead to improbable configurations of the deformation field, for example by rejecting samples that move a node in the mesh into an improbable region, or rejecting samples that lead to folding of the elements.

3 Uncertainty Visualization in Non-rigid Registration

From the marginal posteriors on node deformations we can estimate the most likely deformation, as well as the uncertainty of this estimate. The most likely deformation is a simple object and is easily visualized, e.g. in the form of deformed versions of the pre-operative image data. Conversely, the full posterior on deformations is a high-dimensional object and is difficult to visualize. In this section we present; 1) ways of summarizing the uncertainty using robust statistics and, 2) how these low-dimensional uncertainty summaries can be visualized.

3.1 Inter-Quartile Ranges

The difference between the third and first quartile of a distribution, often called the inter-quartile range (IQR), is a robust statistic that conveys the dispersion of a distribution. It is robust in the sense that it provides meaningful information even for non-Gaussian distributions. We propose to use IQRs of the marginal

distributions as a measure of uncertainty and visualize them, either as ellipsoids, or as a scalar map:

IQR Ellipsoids. At each node in the FE-mesh we compute the IQRs of the x, y and z samples and fit an ellipsoid to these ranges. An IQR ellipsoid provides a 3D view of the extent of the marginal distribution on deformations at this node.

Maximum IQR Scalar Map. Sometimes it is easier to interpret a scalar measure of the statistical dispersion. A maximum IQR scalar map can be generated by visualizing the maximum of the three IQRs (there is one IQR for each dimension) at a point. This measure, however, is not able to convey any bias towards dispersion in a single dimension.

3.2 Summarizing and Visualization of Marginal Distributions

In neurosurgery, a non-rigid mapping between the pre- and intra-operative image spaces is established mainly to update the intra-operative space with functional information from the pre-operative space. The functional information is commonly in the form of either a surface model outlining eloquent functional areas extracted from functional Magnetic Resonance Imaging (fMRI), or streamlines describing fiber tracts extracted from Diffusion Tensor Imaging (DTI). For each deformation sample, we deform both the surface models and line data into the intra-operative space. If we draw N samples to characterize the posterior, then we have N deformed versions of the pre-operative models in the intra-operative space. Depending on whether we have surface models (fMRI) or line data (DTI), we apply different techniques for reducing the high-dimensional data to a low-dimensional space that is more suitable for visualization:

fMRI: Marginal Volume. We construct a marginal integer image volume, or a histogram volume [9], such that for each deformed surface model (fMRI activated volume) we increment the value at each voxel that is inside the surface model and divide by N. This is the probability for a voxel to be in the fMRI activation volume.

DTI: Marginal Visitation Volume. For the deformed fiber tracts we use a technique called visitation count [6]. A visitation count volume is constructed by tracing the deformed tractography streamlines through the volume and incrementing values in each voxel a streamline crosses. This is the marginal distribution that a fiber crosses a voxel.

Marginal Confidence Bounds. Given a marginal volume, we can extract isocontours and thereby render marginal confidence bounds. For instance, the 25% confidence bound is the contour where voxels outside the bound were inside the deformed model in less than 25% of the samples.

The marginal volume and the marginal visitation volume can be visualized either using volume rendering to provide a 3D view of the uncertainty, or with a colormap that conveys the marginal density overlaid on a 2D slice of the anatomical data.

4 Results

From registration of a clinical neurosurgical dataset, we present results in the form of visualizations of the registration uncertainty. The dataset consists of pre-operative anatomical MRI, fMRI and DTI, and a post-operative anatomical MRI used as a proxy for an intra-operative image of a patient with a large glioma lesion in the left frontal region near speech areas.

4.1 Dataset

The anatomical images consisted of a pre-operative 512x512x144 T1 MR image with spacing (0.5mm, 0.5mm, 1.0mm) and a post-operative 256x256x130 T1 MR image with spacing (0.94mm, 0.94mm, 1.4mm) acquired two days post surgery. We extracted tractography streamlines from DTI images and an iso-surface from an fMRI activation volume of the language center located closest to the tumor. A FE mesh covering the brain was constructed with 933 nodes and 4262 tetrahedral elements. The dataset can be seen in Fig. 1. Notice that the post-operative image is contrast-enhanced which makes it difficult to identify the size of the resection due to the bright appearance of blood and edema around the resected area.

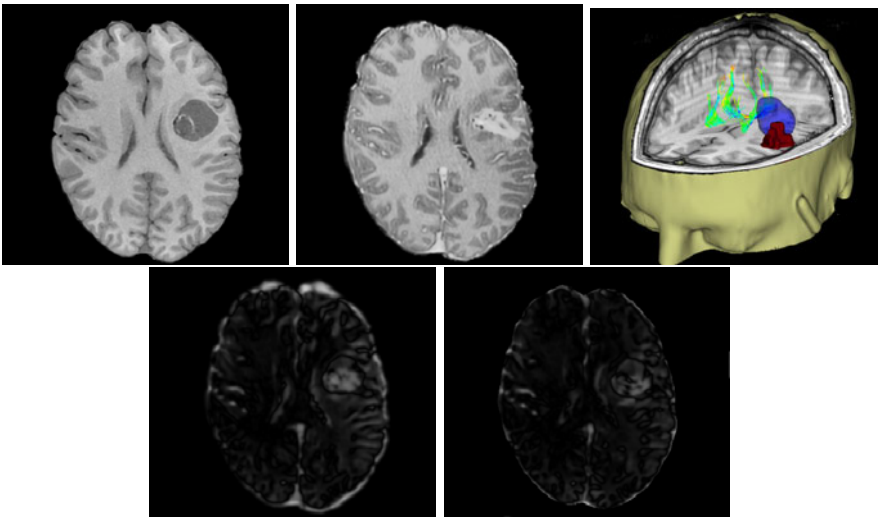


Fig. 1. [a b c; d e]: Dataset. (a)-(b) Pre- and post-operative image. (c) A 3D composite view of the pre-operative volume augmented with the tumor (blue), speech area (red) and the diffusion fiber tracts. (d) Absolute difference image between the two corresponding slices in (a)-(b) before registration. (e) Absolute difference image after registration. Notice the large intensity disagreement around the tumor because of edema and blood. Also notice that, due to contrast enhancement, blood vessels are bright in the post-operative image but dark in the pre-operative image which might lead to an increase in registration uncertainty.

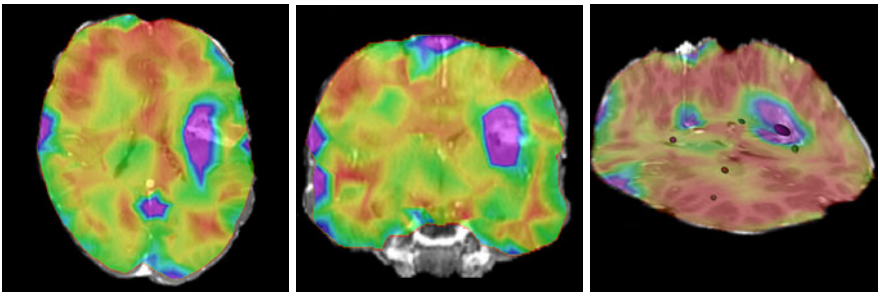


Fig. 2. [a, b, c]: Inter-quartile visualizations overlaid on the post-operative image. (a)-(b) Axial and coronal maximum IQR maps. Pink areas have maximum IQR of 7.5 mm which decreases down to 1.7 mm for red. Notice the large maximum IQRs around the tumor. We suspect the reason is twofold: 1) the large intensity difference around the tumor due to blood and edema and 2) after resection there is no longer a one-to-one correspondence between the pre- and post-operative image. (c) Composite visualization of intersecting axial, coronal and sagittal planes overlaid with the maximum IQR map and a few IQR ellipsoids. Notice the elongation of the ellipsoid at the tumor.

4.2 Deformation Parameter Estimation

To speed up computations, three parallel MCMC samplers, each generating approximately 11 samples per second, were used. We assumed the chains had mixed, and samples could be pooled together, when the *potential scale reduction* [8] dropped below 1.2 for each scalar estimand. The elastic parameters were set to $\lambda = 25,\ \mu = 1$, and the temperatures and the size of the jumping kernels ($T_s = 1.5,\ T_r = 800$ and $\sigma_t = 0.03$) were set to achieve an acceptance rate of approximately 25%. We let each sampler generate 300 000 samples, discarded the first 100 000 samples, used a thinning factor of 10 and pooled together the remaining samples for a total of $N = 60\ 000$ samples that were used for the estimations. The difference image between the pre- and intra-operative image after registration using the most likely deformation can be seen in Fig. 1(e).

4.3 Uncertainty Visualization

In Fig. 2 we visualize IQRs computed from the marginal posteriors in the form of scalar maps and ellipsoids. Notice that the uncertainty is large around the tumor site. The uncertainty, as well as the most likely location of the deformed functional information, is visualized in Fig. 3 using marginal volumes, marginal visitation count volumes and marginal confidence bounds. In Fig. 3(a) we see that the most likely boundary of the activated functional area is far from the area of resection, however, the probability map in Fig. 3(b) shows that there is some probability that parts the functional area might have been touched during surgery. This suggests the importance of uncertainty visualization in surgical decision making.

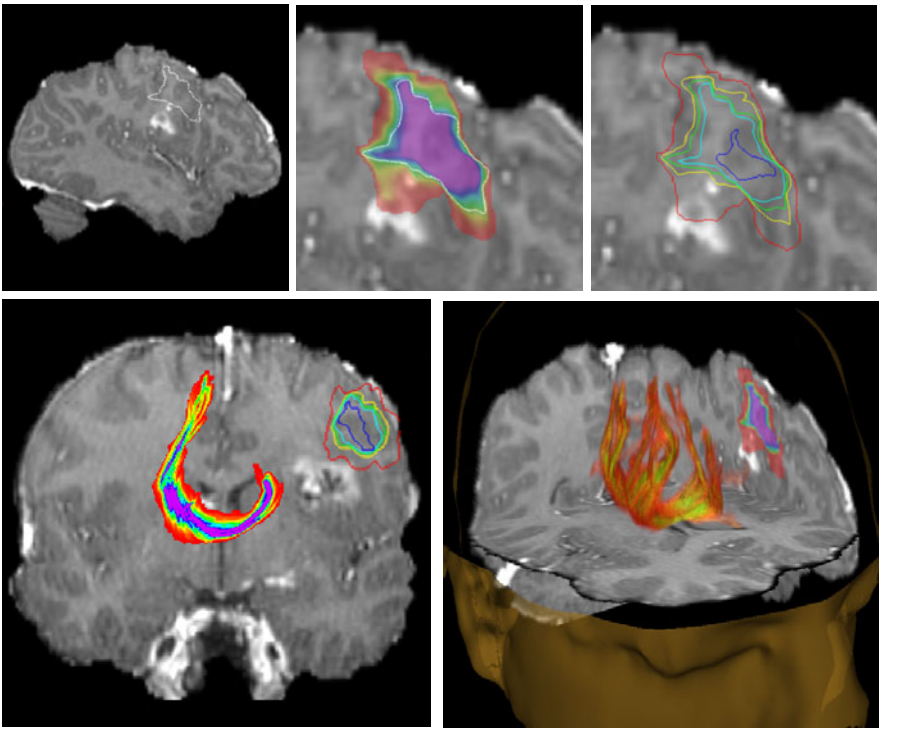


Fig. 3. [a, b, c; d, e]: Marginal volumes. (a) Sagittal slice with the most likely estimate of the fMRI activated area outlined with a white contour. (b)-(c) Close up of (a). In (b) we render the marginal fMRI activation volume using a colormap. The probability density decreases across the iso-contours from purple to red. (c) Rendering of the marginal quartile bounds. The most likely estimate in (a) is close to the 50% marginal quartile bound. (d) A coronal slice overlaid with marginal confidence bounds of the fMRI activation area as well as the density colormap of the marginal visitation count volume. (e) A composite view of three orthogonal planes, with volume rendering of the marginal visitation count volume and the density of the marginal activation volume as a colormap. In the conventional approach to registration, in which only the optimal deformation is presented, the display would only show (a), which would give the impression that the resection did not touch on the fMRI activated area (the high intensity area is due to blood and edema caused by the resection). However, if we take the uncertainty information in (b) into account, we may conclude that there is moderate probability that the surgeon touched the fMRI activated area. This additional information might have significant impact on surgical decision making.

5 Discussion

While the specific registration uncertainty presented in the examples is clearly a preliminary result, we believe, firstly, that in the future, with increased computational power, better modeling, and more highly developed algorithms, it will be practical to obtain dense information about the uncertainty of the estimated registration, and secondly that it will be important to share that information with the surgeon in a meaningful way. Our goal in this paper has been to explore the latter issue, and to promote the importance of these issues to the research community. We introduced several visualization methods to convey registration uncertainty, and the experiments show that it can be important to convey this information, especially for functional areas located close to tumors. Some comments from neurosurgeons that were shown the uncertainty visualizations were: 1) this is really important information that should be integrated in their surgical guidance systems and 2) the slice renderings of the marginal volumes are very informative (e.g. Fig. 3(b)) because they convey information similar to the original fMRI activation volumes which clinicians know how to interpret. Future work should include the functional uncertainty in the registration model.

Acknowledgments. We are grateful to M.D. Alex Golby for her comments. This work was supported by NIH grants R01CA138419, P41RR13218, U54EB005149 and U41RR019703.

References

1. Hub, M., Kessler, M., Karger, C.: A stochastic approach to estimate the uncertainty-involved in b-spline image registration. IEEE Trans. on Med. Im. 28(11), 1708–1716 (2009)
2. Kybic, J.: Bootstrap resampling for image registration uncertainty estimation without ground truth. IEEE Trans. on Im. Proc. 19(1), 64–73 (2010)
3. Kybic, J.: Fast no ground truth image registration accuracy evaluation: Comparison of bootstrap and hessian approaches. In: ISBI, pp. 792–795 (2008)
4. Simpson, A.L., Ma, B., Chen, E.C.S., Ellis, R.E., Stewart, A.J.: Using registration uncertainty visualization in a user study of a simple surgical task. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 397–404. Springer, Heidelberg (2006)
5. McCormick, T., Dink, D., Orcun, S., Pekny, J., Rardin, R., Baxter, L., Thai, V., Langer, M.: Target volume uncertainty and a method to visualize its effect on the target dose prescription. Int. J. of Rad. Onc. 60(5), 1580–1588 (2004)
6. Jones, D.K., Pierpaoli, C.: Confidence mapping in diffusion tensor magnetic resonance imaging tractography using a bootstrap approach. Magn. Res. in Med. 53(5), 1143–1149 (2005)
7. Risholm, P., Samset, E., Wells, W.M.: Bayesian estimation of deformation and elastic parameters in non-rigid registration. In: Workshop on Biomedical Image Registration (to appear, 2010)
8. Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B.: Bayesian Data Analysis, 2nd edn. Texts in Statistical Science. Chapman & Hall/CRC (July 2003)
9. Fan, A.C., Fisher, J., Wells, W.M., Levitt, J., Willsky, A.: MCMC curve sampling for image segmentation. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 477–485. Springer, Heidelberg (2007)

Coupled Registration-Segmentation: Application to Femur Analysis with Intra-subject Multiple Levels of Detail MRI Data

Jérôme Schmid*, Jinman Kim, and Nadia Magnenat-Thalmann

MIRALab, University of Geneva, Switzerland
{schmid,jinman.kim,thalmann}@miralab.ch

Abstract. The acquisition of intra-subject data from multiple images is routinely performed to provide complementary information where a single image is not sufficient. However, these images are not always co-registered since they are acquired with different scanners, affected by subject's movements during scans, and consist of different image attributes, e.g. image resolution, field of view (FOV) and intensity distributions. In this study, we propose a coupled registration-segmentation framework that simultaneously registers and segments intra-subject images with different image attributes. The proposed coupled framework is demonstrated with the processing of multiple level of detail (LOD) MRI acquisitions of the hip joint structures, which yield efficient and automated approaches to analyze soft tissues (from high-resolution MRI) in conjunction with the entire hip joint structures (from low resolution MRI).

Keywords: Segmentation; Registration; Deformable models; Hip joint; Level of detail.

1 Introduction

The acquisition of intra-subject data from multiple images is routinely performed to provide complementary information where a single image is not sufficient. As an example, the acquisition of MRI images of the hip joint at multiple levels of detail (LOD) is often used to identify bone changes and soft tissue abnormalities, which might explain the development of arthritis. In this protocol, exemplified in Fig. 2, a low resolution MRI that covers a large field of view (FOV) is acquired to image the whole hip joint structures (i.e. entire femur bone), whereas a high resolution MRI with limited FOV and centered at the hip joint is acquired to reveal the fine details of the soft tissues. Similarly, the use of multiple images of joints acquired in different postures was demonstrated to be an effective approach for biomechanical motion analysis, which required these images to be aligned (e.g., [1]). From these examples, there is a crucial need for image registration where multiple acquisitions are not always co-registered. In addition, the segmentation

* Corresponding author.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 562–569, 2010.

of the structures of interest (SOIs), shared among the acquisitions, is commonly required for quantitative correspondence.

The ability for concurrent segmentation and registration has been demonstrated to be a powerful method in recent years. A seminal paper by Yezzi et al. [2] presented a variational framework for simultaneous segmentation and registration using deformable models, and gave birth to a variety of methods [3,4,5,6,7] with extensions such as the support for more than 2 images or non-rigid transforms. In other studies, voxel-based approaches, formulated as a maximum a posteriori (MAP) estimation problem were proposed [8,9]. Atlas-based methods [5,9] were also presented, but were limited to data with atlas availability. Further, atlas-based methods generally involved non-rigid transforms, and were thus not appropriate to e.g. a *rigid* structure imaged in different positions.

In this paper, we propose a coupled registration-segmentation framework based on deformable models for intra-subject images acquired with different protocols. The framework is able to handle intra-subject images characterized by varying FOV in which SOIs can be *partially visible* and thus have a limited correspondence which seriously complicates their direct registration. Further, our framework is atlas-free, does not limit the number of images or SOIs, and does not restrict the nature or evolution strategies of the deformable models. Its effectiveness is demonstrated in the registration-segmentation of femurs in intra-subject MRIs acquired at multiple LODs (hereon referred to as "multiple LOD MRIs"). Our coupled framework will simultaneously segment and register the multiple LOD MRIs, and thus enable an efficient and automated approach to the analysis of (pathological) soft tissues from images of the joint (high-resolution) in conjunction with the entire hip structure (low-resolution).

2 Coupled Registration-Segmentation

2.1 Framework Overview

In this work we propose to segment and register the same SOIs in N images $I^1, \ldots, I^N$ $(I^i : \Omega^i \rightarrow \Re, \Omega^i \in \Re^D (D = 2, 3))$. For the sake of brevity, we only present our methodology in case of a single SOI, while the proposed approach can be easily extended to multiple SOIs. We assume that the segmentation is achieved by using N deformable models (DM) represented by shapes $S^1, \ldots, S^N$, composed of a same number n of points x_j^i: $S^i = \{x_1^i, \ldots, x_n^i\}$. No restrictions are made on the nature of the DM (e.g., curves [4], simplex meshes [10]), except that their shapes must share the same number of points and be in point correspondence. Similarly, DMs can adopt any type of evolution strategies (e.g., Gradient flow [2], Newtonian law equations [11]) as long as they can be expressed as an update operator $\Psi(S^i)$, which returns new state of the shape at each evolution step.

The registration consists in computing the mappings $g^{i \leftarrow j}$ to transform S^j to S^i. Similarly to [7], instead of computing the N^2 possible mappings $g^{i \leftarrow j}$, we estimate the mappings g^i that transform a common shape $\bar{S}$ to each shape S^i. In

case of linear transforms and shapes in point correspondence, we propose to express these mappings g^i by using a generalized Procrustes alignment (GPA) [12]. During this procedure, mean shape $\bar{S} = \{\bar{x_1}, \ldots, \bar{x_n}\}$ and mapping g^i estimates are iteratively updated. The mappings are estimated in a least square sense:

$$g^i = \underset{g}{\operatorname{argmin}} \sum_j \left\| x_j^i - g(\bar{x_j}) \right\|^2 = \underset{g}{\operatorname{argmin}} \sum_j d_g(x_j^i)^2 \ . \tag{1}$$

An overview of the steps involved in our coupled registration-segmentation framework is depicted in Fig. 1, in which segmentation and registration are used in an interleaved manner. The steps are repeated until segmentations convergence. First shapes S^i are updated given the segmentation update operator Ψ. Then the resulting shapes $\Psi(S^i)$ are aligned with the GPA. The final stage performs a linear "blending" at the point level between shapes $\Psi(S^i)$ and the back-transformed mean shapes $g^i(\bar{S})$ to get new shapes S'^i:

$$S'^i = \lambda g^i(\bar{S}) + (1 - \lambda)\Psi(S^i) \ . \tag{2}$$

Parameter $\lambda \in [0, 1]$ is used as a stiffness coefficient, which expresses the degree of constraint applied by the registration to the segmentation. Theoretically, if shapes only differ by a linear transform λ should be set to 1. However, in practice it is better to slightly relax λ at the beginning of the segmentation. This provides more freedom to the segmentation, especially when the shapes are not initialized closely enough and they need to be significantly deformed. In this paper, we linearly increases λ from 0.7 to 1.

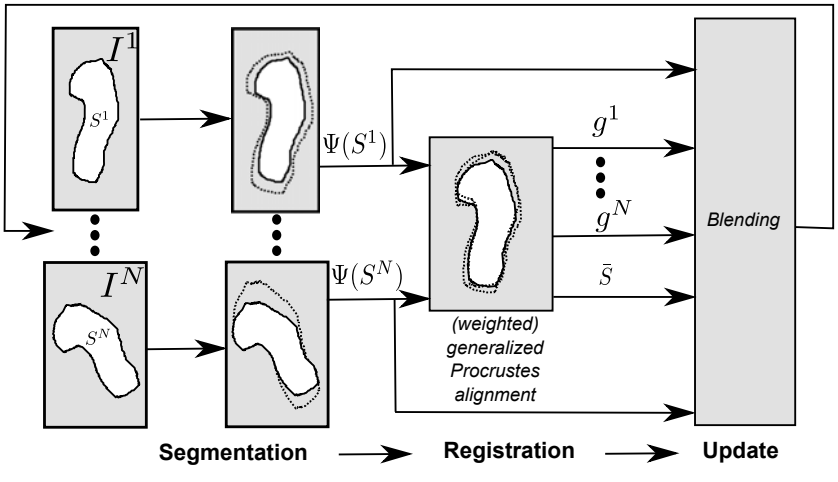


Fig. 1. Coupled registration-segmentation step: initially, a segmentation step is executed, evolving N shapes $S^i \rightarrow \Psi(S^i)$; resulting shapes $\Psi(S^i)$ are then aligned with a (weighted) GPA to yield alignment mappings g^i and mean shape $\bar{S}$); finally, shapes $\Psi(S^i)$ and back-transformed mean shapes $g^i(\bar{S})$ are blended to produce new shapes

2.2 Weighting Procedure

A major issue with the aforementioned step is that all shape points are equally treated in the registration. If these points are invalid from a segmentation viewpoint (e.g. point not lying on organ boundary or out of the image FOV), they will corrupt the registration. As a result, bad estimates of the mean $\bar{S}$ and mappings g^i will be computed. Hence we apply a simple yet efficient solution that uses a *weighted* GPA. Given weights w_j^i associated with point x_j^i of shape S^i, the weighted alignment is obtained by weighting (i) the mean (i.e. $\bar{x_j} = \sum_i w_j^i x_j^i / \sum_i w_j^i$) and (ii) the least square minimization in Eq. (1), by replacing the distance terms $d_g(x_j^i)$ with $w_j^i d_g(x_j^i)$.

In this paper, we express the weight w_j^i as the sum of "reliability" terms $F^k(\theta(x_j^i))$: $w_j^i = \sum_{1 \leq k \leq N} F^k(\theta(x_j^i))$. The quantity $\theta(x_j^i) = \theta_j^i$ is a features vector computed from the image I^i, the shape S^i and the point x_j^i. The *intra*-reliability term $F^i(\theta_j^i)$ accounts for the segmentation quality at x_j^i in image I^i. Similarly, *inter*-reliability terms $F^k(\theta_j^i), k \neq i$ additionally consider the coupling with other images. These terms depend on the segmentation strategy and available prior knowledge (see next Section for examples). If we consider (i) the features vector θ_j^i as the observation of a random variable Θ, (ii) the event w "θ_j^i is the features vector obtained in case of optimal segmentation at x_j^i", and (iii) $F^k(\theta_j^i) = p^k(w|\theta_j^i)$ as a posteriori probability function, w_j^i computation is analog to a "sum rule" commonly used in multi-classifiers approaches.

3 Coupled MRI Bone Registration-Segmentation

Our experiments are related to the hip joint image analysis to detect abnormal bone shape (changes) that may yield the formation of arthritis [13]. We further suggest that the automated alignment of multiple LOD MRIs, together with its segmentation will better support clinical diagnosis and biomechanical studies. We adopted our previous MRI bone segmentation approach [11], which uses DMs that are semi-automatically initialized and driven by dynamic law equations. The evolution is coupled with shape priors expressed as multi-resolution statistical shape models (SSM). Convergence is obtained when shapes variations between two evolution steps are small with respect to a chosen threshold. Image forces used in the DM evolution are based on gradient directions and on the normalized cross correlation ratio $NCC(q_j^i, rq_j^i)$ between intensity profiles [10] q_j^i extracted at each point x_j^i and reference profiles rq_j^i. When the segmentation is near completion, the effects of shape priors are turned off enabling the segmentation of fine details not captured by the SSM.

Two acquisition protocols VIBE (VB) and TrueFISP (TF) were used to get the low ($1.3 \times 1.3 \times 5$ mm^3, Fig. 2a) and high ($0.6 \times 0.52 \times 0.52$ mm^3, Fig. 2b) resolution MRIs, respectively. As depicted in Fig. 2, their FOV differ significantly where the TF focuses on the joint while the VB covers the full femur bone. The computation of weights of Sec. 2.2 is critical. Weights are constantly updated

during the segmentation as follows: shape points out of their corresponding image FOV have null weights (lowest weight), otherwise the NCC ratio is used to define the intra-reliability $F^i(\theta_j^i)$ and inter-reliability $F^k(\theta_j^i)$ as $NCC(q_j^i, rq_j^i)$ and $NCC(q_j^i, q_j^k)$, respectively. By giving higher weights to a shape point whose IP is similar to the other shapes' IPs, robust estimates of the mean shape and alignment transforms are obtained. This provides thus an effective way to express point reliability. The NCC is invariant to linear intensity changes, thus being useful for our MRIs acquired with different protocols.

4 Results

Fifteen volunteers were scanned based on Sec. 3 imaging protocols. For each subject, multiple LOD MRIs (TF and VB) were processed with our proposed coupled registration-segmentation framework. For comparative analysis, the VB was also segmented using the same deformable model as in the coupled approach. This is referred to as a single segmentation. The VB image was segmented instead of the TF in order to have a more reliable and objective estimate since in [11] images with small FOV were not used. For quantitative analysis, the Dice's similarity coefficient (DSC) was used to compare the automated femur segmentation results with that of manual delineations (as the ground truths) defined by experienced researchers under the supervision of a radiologist. A DSC of 1.0 was given to segmentation results having identical overlap to the ground truth. In the first experiment, multiple LODs of VB and TF were used, in which the

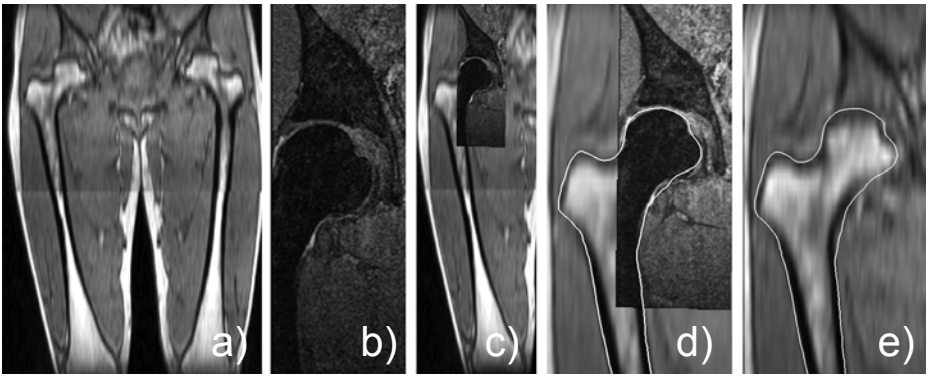


Fig. 2. a) VB and b) TF images of a same subject acquired at different LODs and positions. The higher LOD of the TF is clearly able to image the cartilages and the femoral head in greater details than with the VB. On the other hand, VB provides full imaging of the femur. These multiple LOD images offer a unique possibility to simultaneously analyze in correspondence, the full femur together with the fine detailed soft tissue structures. The alignment resulting from our coupled approach is shown in c) with corresponding *white* segmentation contours overlaid in d). In comparison, single segmentation results of a) are shown in e).

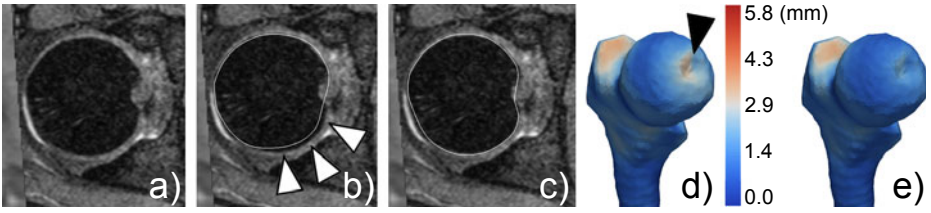


Fig. 3. a) A close-up transversal view of a femoral head visible in the TF. Results from b) single and c) coupled segmentation are overlaid as *white contours*. The *white arrows* indicate poorly segmented areas in b) which are markedly corrected in c) using the additional information from the TF MRI. Average distance errors computed from the single d) and coupled e) segmentation results are mapped on an arbitrary femur shape. The *black arrow* indicates the Fovea Capitis for which larger error differences are observed.

subjects' leg exhibited low flexion/abduction/adduction amplitudes (i.e. close to the "neutral" position) as shown in Fig. 2a) and b).

The result of our coupled approach is shown in Fig. 2c) and d), where both LOD MRIs have been automatically aligned (superimposed together) and segmented, respectively. Here, we can see that the alignment appears to be satisfactory and the segmentation shape correctly contours the bone in both VB and TF MRIs. The surrounding soft tissues, i.e. muscles, were slightly misaligned, although visibly accurate. This was expected since the alignment was based on a rigid mapping of the femur bone and thus did not take into account the soft tissue deformations. In Fig. 2e), the segmentation result using only the VB is presented. Visual differences to the coupled approach results are very subtle. The visual findings are validated with the average DSC measures of 0.917 ± 0.03 and 0.925 ± 0.01 for the single and coupled segmentations, respectively.

Although the improvements with the coupled approach are small for the entire femur segmentation ($\approx$1% increase in DSC), the coupled approach is markedly more accurate in the segmentation of the finer details that were only available in the TF (Fig. 3). A distance error (in mm) was calculated between the ground truth and the segmentation results for both the single (Fig. 3d) and the coupled (Fig. 3e) approaches, averaged among all the subjects, and finally mapped on an arbitrary reference shape. These distance errors better reveal that the finer information from the TF contributed to a better segmentation (e.g., fine details of the fovea capitis). In this first experiment, the position of the leg was neutral for both the VB and the TF MRIs. However, in clinical practice, it often happens that stronger alignment differences between the image acquisitions are observed due to patient movement between the scans. Large leg rotation patterns make the alignment more challenging and locally affect the intensity distribution around the bones. In order to analyze the capability of our coupled framework in the presence of large movement between the scans, we gathered 15 additional datasets of the same subjects performing large leg rotation patterns during the TF acquisition, as shown in Fig. 4.

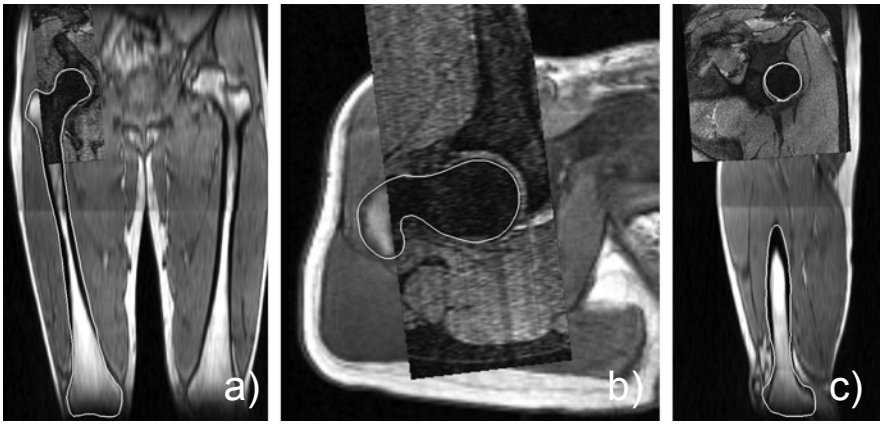


Fig. 4. Segmentation and alignment result of a multiple LOD MRIs of VB and TF shown in a) coronal b) transversal and c) sagittal view slices. In this example, the subject performed a large rotation movement of the leg.

These figures illustrate the strong misalignment that needed to be recovered between the bone extracted in the VB and the TF MRIs. As shown in Fig. 4, the registration of the femur performed well, where we can clearly see the shape outline of the femur between the VB and the TF. Similarly to the previous experiments (Fig. 2), the quantitative measures of the femur segmentation were almost identical (DSC of 0.917±0.03 and 0.924±0.01 for single and coupled approaches respectively). This suggests that our framework is robust in regards to the mis-alignment orientation of the LOD MRIs as long as these MRIs share a common SOI. We conducted further tests to evaluate our coupled approach in the case where more than two LOD MRIs were available. When our framework was applied to multiple LOD MRIs consisting of a VB and two TF (neutral and large movement), all three images were successfully aligned and the segmentation accuracy was in line with all the other experiments (alone and coupled DSC results of 0.917±0.03 and 0.929±0.01, respectively).

5 Discussion and Conclusion

This study presented a framework for simultaneous registration and segmentation of intra-subject multi-modal images. We exemplified our framework with the application to multiple LOD MRIs. The experimental results demonstrated that our framework was able to robustly align the multiple LOD MRIs that were different in resolution and FOV. At the same time, the combined information in the multiple LOD MRIs was exploited in the segmentation, thus improving the results when compared to the segmentation using a single MRI. We measured our segmentation results to those derived from manual delineation and found the results to be highly accurate. Since the accuracy of the registration is dependent

on the segmentation in our coupled approach, we suggest that our registration is equally accurate. With our coupled registration-segmentation framework, high and low resolution images of the hip joint structures were simultaneously aligned and segmented, thus offering quantitative correspondence for use in e.g. clinical diagnosis and biomechanical analysis. Current effort is put in automating the initialization of the DMs. Our future work will involve the evaluation of our coupled framework to a wider variety of imaging modalities as well as the extension to non-rigid transforms.

Acknowledgments. This work was supported by the 3D Anatomical Human Project (MRTN-CT-2006-035763) funded by the European Union. We thank our partners, the University Hospital of Geneva and the Great Theater of Geneva.

References

1. McPherson, A., Kärrholm, J., Pinskerova, V., Sosna, A., Martelli, S.: Imaging knee position using MRI, RSA/CT and 3D digitisation. J. Biomech. 38(2), 263–268 (2005)
2. Yezzi, A., Zollei, L., Kapur, T.: A variational framework for joint segmentation and registration. In: Proc. IEEE MMBIA, pp. 44–51 (2001)
3. Moelich, M., Chan, T.: Joint segmentation and registration using logic models. Technical Report CAM Report 03-06, UCLA (2003)
4. Unal, G., Slabaugh, G.: Coupled PDEs for non-rigid registration and segmentation. In: Proc. CVPR, vol. 1, p. 168. IEEE Computer Society Press, Los Alamitos (2005)
5. Wang, F., Vemuri, B.C.: Simultaneous registration and segmentation of anatomical structures from brain MRI. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 17–25. Springer, Heidelberg (2005)
6. Droske, M., Rumpf, M.: Multiscale joint segmentation and registration of image morphology. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(12), 2181–2194 (2007)
7. Lord, N., Ho, J., Vemuri, B.: USSR: A unified framework for simultaneous smoothing, segmentation, and registration of multiple images. In: Proc. ICCV, pp. 1–6 (2007)
8. Wyatt, P., Noble, J.: MAP MRF joint segmentation and registration of medical images. Med. Image Anal. 7(4), 539–552 (2003)
9. Pohl, K.M., Fisher, J.W., Levitt, J.J., Shenton, M.E., Kikinis, R., Grimson, W.E.L., Wells, W.M.: A Unifying Approach to Registration, Segmentation, and Intensity Correction. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 310–318. Springer, Heidelberg (2005)
10. Delingette, H.: General Object Reconstruction Based on Simplex Meshes. Int. J. Comput. Vis. 32(2), 111–146 (1999)
11. Schmid, J., Magnenat-Thalmann, N.: MRI Bone Segmentation using Deformable Models and Shape Priors. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part I. LNCS, vol. 5241, pp. 119–126. Springer, Heidelberg (2008)
12. Gower, J.: Generalized Procrustes analysis. Psychometrika 40, 33–51 (1975)
13. Evangelisto, A., Wakefield, R., Emery, P.: Imaging in early arthritis. Best Pract. Res. Clin. Rheumatol. 18(6), 927–943 (2004)

Groupwise Registration with Sharp Mean

Guorong Wu[1], Hongjun Jia[1], Qian Wang[1,2], and Dinggang Shen[1]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
{grwu,jiahj,dgshen}@med.unc.edu
[2] Department of Computer Science, University of North Carolina at Chapel Hill
qianwang@cs.unc.edu

Abstract. Groupwise registration has received more and more attention in the area of medical image analysis, due to its importance in analysis of population data. One popular way for groupwise registration is to alternatively estimate the group mean image and register all subject images to the estimated group mean. However, for achieving better registration performance, it is important to always keep the sharpness of the group mean image during the registration, which has not been well investigated yet in the literature. To achieve this, we propose to treat each aligned subject, as well as its anatomical regions, differently when constructing the group mean image. Specifically, we propose a new objective function to generalize the conventional groupwise registration method by using a dynamic weighting strategy to weight adaptively across subjects and spatial regions, to construct a sharp group mean image in each stage of registration. By integrating this strategy into diffeomorphic demons algorithm, the performance of our groupwise registration can be significantly improved, compared to the conventional groupwise registration method that starts with a fuzzy group mean image.

1 Introduction

Groupwise registration becomes more and more popular in recent years due to its attractiveness in analyzing population data [1, 2]. Compared to the traditional pairwise registration algorithm, groupwise registration aims to simultaneously estimate the transformation fields for all subjects without explicitly specifying an individual subject as a template, in order to avoid any bias in the subsequent data analysis.

One of the most popular groupwise registrations is proposed by Joshi *et al.* [1] in 2004. In their method, the groupwise registration is implemented by alternatively constructing the group mean image and estimating the transformation fields of all subjects towards the tentative group mean. However, one major drawback of this method is that it usually produces a very fuzzy group mean image in the beginning, due to the simple averaging of all subjects that are not well aligned initially. As a result, the fuzzy group mean image fails to offer the correct guidance to the subsequent pairwise registrations as we can see in the experiments. Thus the anatomical details can hardly be recovered from the initial fuzzy group mean image due to the difficulty of establishing the reliable correspondences between sharp subject images and the fuzzy group mean image during the iterative registration procedure.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 570–577, 2010.

Recently, Fletcher *et al.* [2] extended Joshi's method to the Riemannian manifold and proposed to use the geometric mean of the group to handle the possible outliers which may deviate the Fréchet mean far away from the real population center. However, they used an equal weight for each subject in the group to compute the geometric mean. Also, the importance of the sharpness of the mean image in groupwise registration is not addressed.

Another interesting related work is the tree-based groupwise registration [3]. In their method, the pseudo-geodesic median image is selected as the root template after learning the intrinsic manifold on the whole data set. Since a fixed image (i.e., the root image) is used as the final template to register all other subjects, the bias is inevitably introduced in this scenario due to the discrepancy between the template and the real population center, although in this case a sharp individual image is used as the template to guide the groupwise registration.

We propose to improve the performance of groupwise registration methods that focus on iteratively registering subjects to the group mean image [1, 3, 4]. We will *first* point out the importance of always keeping a sharp mean image during the groupwise registration. *Second,* we generalize the conventional groupwise registration method by presenting a new objective function, for achieving the sharp group mean image without introducing biases. To accomplish it, we learn the distance between the two images on local patch of the manifold, instead of the straightforward but coarse measurement defined as the overall intensity difference of the whole brain [1-3]. Another contribution of this paper is that we treat each subject differently throughout the registration. Specifically, only the registered subjects that are close enough to the tentative mean image will be involved in updating the mean image since equally treating the subjects in the early stage will lead to the irreversible loss of structure details, especially when most subjects are not well aligned in the beginning. With the improvement of registration, subjects are more likely to agglomerate to the population center. At that time, more subjects will be allowed to participate into the construction of the mean image and their contributions will become more equal. Working under this scenario, the group mean image in our method will gradually approach to the population center, as shown by our experimental results. *Finally*, we demonstrate the advantage of our method by integrating it into the Diffeomorphic Demons registration algorithm [5], and also compare its performance with the conventional groupwise registration algorithm [1]. With extensive experiments in evaluation of the overlap ratios on 16 NIREP Data and 40 LONI data, our proposed method outperforms the conventional groupwise registration method in both registration accuracy and consistency.

2 Methods

In the framework of unbiased groupwise registration [1], the transformation fields are estimated by iteratively registering N subjects to the latest estimated group mean image. In the t-th round of registration ($t = 1, \dots, T$), the group mean image $\boldsymbol{M}^t$ is generated by averaging upon the warped subjects $\boldsymbol{I}^t = \{\boldsymbol{I}_s^t | s = 1, \dots, N, t = 1, \dots, T\}$ w.r.t. the current estimated transformation fields $\boldsymbol{G}^t = \{g_s^t | s = 1, \dots, N, t = 1, \dots, T\}$, where $\boldsymbol{G}^t$ are calculated by considering M^{t-1} as the template and $\boldsymbol{I}^0$ (original subjects) as the moving images. It is worth noting that $\boldsymbol{G}^0$ are the linear transforms in the initial

registration stage. In the following, we first point out in Section 2.1 the importance of always keeping the sharp mean image throughout the whole groupwise registration procedure. Then we propose an improved objective function in Section 2.2, and finally optimize it in Section 2.3.

2.1 The Importance of Keeping Sharp Mean Image in Registration

Unbiased groupwise registration method [1] seeks to alternatively estimate the group mean and register each subject to the tentative group mean. However, the initial group mean image M^0, generated right after the linear alignment, is generally very fuzzy, since the subjects are not well aligned in the beginning of registration. According to our knowledge, few articles have addressed the importance of keeping the sharp mean image during the registration. Indeed, the fuzzy mean image would undermine the groupwise registration performance in two ways: 1) it is difficult to register an individual subject with clear anatomical structures to the mean image with fuzzy structures; 2) a fuzzy mean image will challenge the convergence of optimization since it might not provide sufficient anatomical information to guide the registration. The importance of a sharp mean image in groupwise registration is demonstrated in Fig. 1 by 61 toy images with three branches in distribution, with each branch representing one type of cortical folding. For simplicity, only three images are shown in each branch. Before registration, the mean image is very blurry (as shown in Fig. 1(a)). If the groupwise registration starts from this fuzzy mean, it will result in an unsatisfactory mean image (i.e., the one shown in Fig. 1(b)), since the fuzzy mean image in the beginning is not able to informatively guide the whole groupwise registration of individual images.

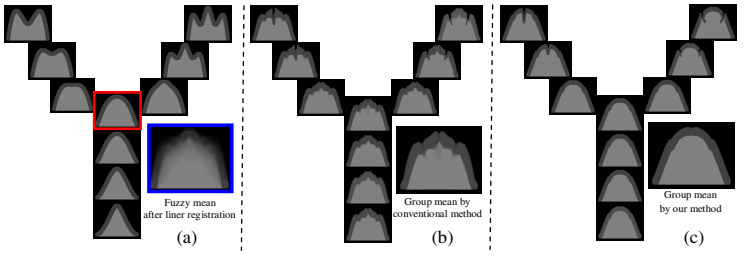


Fig. 1. A toy example to demonstrate the importance of keeping the sharp mean image in the groupwise registration. The synthetic data and their fuzzy mean are shown in (a). (b) shows the groupwise registration results starting with the fuzzy mean image (shown in the blue box of (a)). (c) demonstrates the results by the proposed method which starts with the sharp mean (shown in red box) and keeps tracking the sharp mean image during the registration.

However, if we selected a subject close enough to the population center (i.e., the image in the red box of Fig. 1(a)) as an initial group mean, the anatomical sharpness could be preserved during the registration. Our method is specifically designed to estimate the sharp mean image, which is close to the population center, even in the early stage of the groupwise registration. With this method (detailed in the next section), we can achieve much more reasonable results as shown in Fig. 1 (c).

2.2 Objective Function of Groupwise Registration

In this section, we first present an objective function in our groupwise registration with sharp mean, which generalizes the conventional unbiased groupwise registration algorithm [1].

The objective function in conventional groupwise registration is given as:

$$\{\widehat{\boldsymbol{G}}, \widehat{M}\} = \arg\min_{\boldsymbol{G}^t, M^t} \sum_{s=1}^{N} \left\{ \sum_{x} \left\| g_s^t\left(I_s^0(x)\right) - M^t(x) \right\|^2 + dist(e, g_s^t) \right\}, \tag{1}$$

where $M^t(x) = \frac{1}{N}\sum_{s=1}^{N} I_s^{t-1}(x)$ is the simple average of the warped images after the previous round of registration. The term $dist(e, g_s^t)$ is the manifold distance between the identity transformation e and g_s^t.

One major drawback of this method is that the contributions (or the weights) are the same for *not only* all subjects in the population, *but also* all voxels in each subject. However, different subjects may be aligned at different degrees with the mean image, thus the equal weighting of them could lead to a fuzzy mean image. Also, each anatomical region may be aligned differently with the mean image, therefore the use of the same weight (generally obtained from the entire subject) for each anatomical region may lead to different amounts of fuzziness across different regions of the mean image. Our method attacks this problem in two ways as described next.

First, we propose a distance measurement for each voxel x w.r.t. the current mean image M^t and the warped subject image I_s^t as:

$$D(x, b, I_s^t, M^t) = d\left(\mathcal{P}_b\left(I_s^t(x)\right), \mathcal{P}_b\left(M^t(x)\right)\right), \tag{2}$$

where $\mathcal{P}$ denotes a local image patch around voxel x, with size b. The term d measures the intensity difference between the corresponding local image patches, $\mathcal{P}_b\left(I_s^t(x)\right)$ and $\mathcal{P}_b\left(M^t(x)\right)$, in the images I_s^t and M^t around the voxel x. It is worth noting that D approaches to the global image distance when b is very large, while D becomes voxel-wise difference in Eq. 2 when b is zero. Since the registration results are usually refined from global shape to local shape, the value of b is large in the initial stage of registration and then gradually deceases with the progress of registration.

Second, to treat each subject differently, we introduce a hidden variable $\omega_s^t(x)$ $(s.t.\sum_{s=1}^{N}\omega_s^t(x) = 1)$ to indicate the contribution of each subject I_s^t in the construction of group mean image on a particular voxel x. In the initial stages, all subjects are not well aligned (especially right after linear registration). If we equally weight each subject (i.e., the entropy of the weighting set $\{\omega_s^t(x)|s = 1, \dots, N\}$ is high), it will lead to a fuzzy mean image as demonstrated in Fig. 1. To keep the sharp mean throughout registration, only the warped subjects I_s^t which are close enough to the previous mean image M^{t-1} are qualified to have a large weight of $\omega_s^t(x)$, while other subjects are penalized with small weight $\omega_s^t(x)$. With the progress of registration, all subjects will likely agglomerate to the population center. Thus, all I_s^t, as long as they are close enough to the mean image, will contribute almost equally to the construction of the mean image. Since they are already well aligned, little fuzziness will be brought to the group image. In this paper, the dynamic changes of weights $\{\omega_s^t(x)|s = 1, \dots, N\}$ from strictly binary to loosely uniform are controlled by requiring the entropy of the weighting set $\{\omega_s^t(x)|s = 1, \dots, N\}$ to be increased with the progress of registration.

By replacing the distance measurement with Eq. 2 and adding the dynamic control of weights across subjects to the objective function, we get a new objective function for groupwise registration as:

$$F(\boldsymbol{\omega}^t, M^t, \boldsymbol{G}^t) \propto \sum_x \sum_{s=1}^{N} \{\omega_s^t(x) \cdot D(x, b, I_s^t, M^t) + r \cdot \omega_s^t(x) \cdot \log(\omega_s^t(x))\} + \sum_{s=1}^{N} dist(e, g_s^t). \quad (3)$$

where the scalar r controls the penalty of large distance from I_s^t to M^{t-1}. We will show its important role when explaining the solution to Eq. 3 in Section 2.3.

2.3 Solutions to Our Groupwise Registration

The solution to $\omega_s^t(x)$ in Eq. 3 can be immediately calculated by setting $\partial F / \partial \omega_s^t(x) = 0$:

$$\widehat{\omega}_s^t(x) = \exp\left(-\frac{D(x, b, I_s^t, M^t)}{r}\right). \quad (4)$$

Here r acts as the inverse temperature in the annealing system based on our observations in Section 2.2. Initially, the degree of r is low, i.e., only the subjects that are very close to the population center will be considered to build a new group mean image, in order to keep the mean image sharp. With the progress of groupwise registration, all subjects become closer to the population center. At that stage, the temperature r will be increased to encourage almost equally weighting of all warped subjects I_s^t. It is worth noting that M^0 is the simple average of linearly aligned subjects in the conventional groupwise registration [1], which could be very fuzzy. In our case, when we set the temperature r close to zero in the beginning, M^0 can be computed from only an individual subject that is the closest to the population center, compared to the other subjects in the group. The advantages of this initialization are: 1) it can have a sharp group mean image to allow for better registration with all subjects; 2) it is closer to the final population center than any other subjects. Note that M^0 in our method will not introduce bias since we will gradually increase the value of r to ensure that it reaches the real population center without sacrificing the registration accuracy. This argument is supported by examples in Figs. 2 and 3.

Fig. 2 demonstrates the difference between the conventional and our groupwise registration methods in updating the group mean image, by taking 61 images in Fig.1 as an example. The solid dots (with different colors) in Figs. 2(a) and 2(c) denote for typical images shown in Fig. 1. In the conventional method, M^0 (black cross) is the equal average of all $I_s^{0\prime}$s, which can be easily deviated by the outliers or less registered images (i.e., a faraway dot from the black cross in Fig. 2(a)). Also, during the iterative registration procedure, each image I_s^t is equally considered, regardless of their registration accuracy (as shown within a large circle in the (a)). However, in our method, we first select an individual image, which is the closest to the population center (i.e., red solid dot in the top of (c)), as a group mean image M^0. Another main difference between the conventional and our methods is that in our method only the well aligned subjects will contribute to the construction of the mean image. With the progress of registration, the value r, equal to the size of gray circle in (c), will increase, and thus more and more well aligned images will be considered for construction of the group mean image, thus also avoiding bias in the end. To demonstrate the

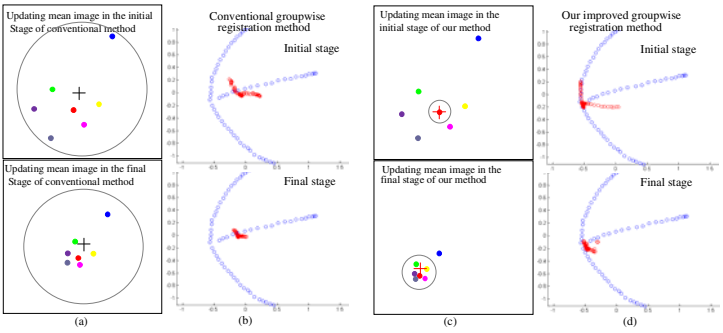


Fig. 2. Different strategies used to update the group mean image in the conventional method (a) and our registration method (c). 61 original subjects used in Fig. 1 are projected onto a 2D space by PCA and shown in blue dots in (b) and (d), respectively. Their registration results by the conventional and our methods are also shown in (b) and (d) with red dots. It can be seen that our registration results are always clustered around the real population center in both initial (top) and final (bottom) registration stages.

advantage of our adaptive strategy in updating the group mean image, we first project 61 images used in Fig. 1 onto a 2D plane by PCA (as shown with blue dots in Figs. 2(b) and 2(d)). The distribution of 61 images shows a three-branch shape, with its center at the joint point of the three branches. The registration results $\boldsymbol{I}^1$ and $\boldsymbol{I}^T$ are also projected onto the same 2D plane, and are displayed in red dots for both the conventional and our methods in (b) and (d), respectively. It can be observed that our registration results always stay around the real population center, while the mean image by the conventional method is deviated from the population center in the beginning and keeps staying there until the end of registration.

After determining the weight $\hat{\omega}_s^t(x)$ for each subject I_s by Eq. 4, the remaining optimization of Eq. 3 will be solved iteratively by updating the group mean image $\hat{M}^t$ and performing pairwise registration between each subject I_s and the group mean image $\hat{M}^t$. Therefore, by fixing g_s^t and $\hat{\omega}_s^t(x)$, we can first calculate the group mean image in the t-th round of registration by minimizing F w.r.t. M^t:

$$\hat{M}^t(x) = \sum_{s=1}^{N} \hat{\omega}_s^t(x) \cdot I_s^t(x) / \sum_{s=1}^{N} \hat{\omega}_s^t(x). \tag{5}$$

Then, the transformation field $\hat{g}_s^t$ in the t-th round of registration can be solved by optimizing the following energy function:

$$\hat{g}_s^t = \arg\min_{g_s^t} \left\{ \sum_x D(x, b, I_s^t, M^t) + dist(e, g_s^t) \right\}, \tag{6}$$

This is a well-known quadratic objective function discussed in many gradient-based registration algorithms. The solution to Eq. 6 can be found in [5].

Compared to the conventional method, the advantage of our objective function can be summarized as: 1) the group mean image is a weighted average of aligned subjects (Eq. 5), instead of a simple equal average. Weights are adaptively determined *not only* for each aligned subject I_s^t, but also for each image location x; 2) the contribution

of each subject is dynamically changed throughout the groupwise registration in the annealing scenario; 3) the sharpness of group mean is always preserved throughout the registration.

3 Experiments

Eighteen elderly brain images are used in our first experiment to demonstrate the advantage of our groupwise registration method, compared to the conventional groupwise registration algorithm [1]. Some typical brain images are displayed in the left panel of Fig. 3. Conventional groupwise registration method starts from a very fuzzy mean (with its 3D rendering shown in blue box). On the contrary, our method begins with a clear group mean image (as shown in red box), which is close to the population center. The evolution of the mean image by the conventional and our registration methods is provided in the top and bottom rows, respectively. It is clear that although the final group mean images are similar by both methods, ours is much sharper than that by the conventional method. Also, this experiment has demonstrated that our method will not introduce bias in the final group mean image as the group mean images by two methods (ours and the conventional unbiased registration algorithm) are very similar. To quantitatively measure the overlap ratio, we first vote a reference image based on the warped results of all (tissue-segmented) subject images, since no template is used in our groupwise registration method. Then, overlap ratio is calculated between each warped subject and the reference image one by one. The average overlap ratio on three tissues (WM, GM, and CSF) is 64.95% by our method and 60.24% by the conventional method, indicating a 4.71% improvement.

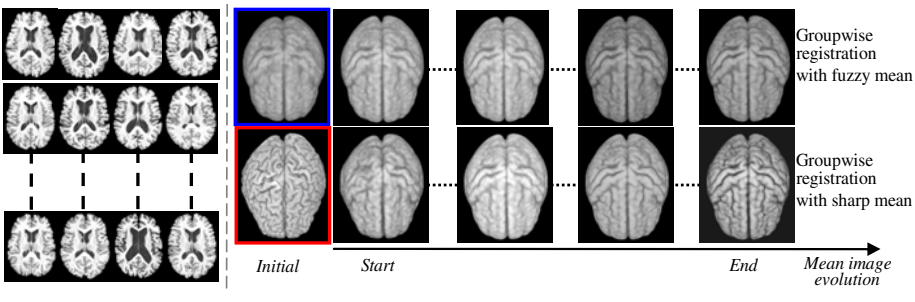
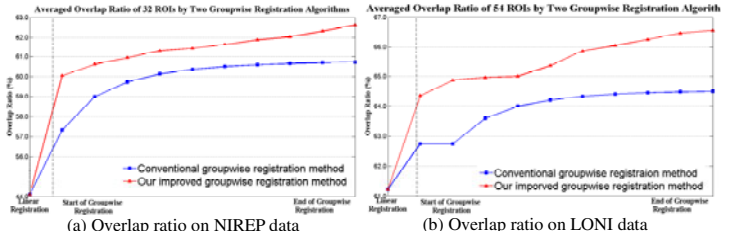


Fig. 3. The evolution of the group mean image. Left panel shows several typical images from 18 elderly brains used in this experiment. In the right panel, the evolution of the conventional groupwise registration starting with a fuzzy mean (in blue box) and our method with a sharp mean (in red box) are displayed in the top and bottom rows, respectively.

To further validate our argument on the importance of using the sharp mean image during the groupwise registration, we perform two different registration approaches, i.e., the conventional groupwise registration with fuzzy mean and our groupwise registration with sharp mean, on 16 NIREP data (with 32 manually labeled ROIs) and 40 LONI data (with 54 ROIs), by measuring the overlap ratios of their aligned ROIs.



(a) Overlap ratio on NIREP data (b) Overlap ratio on LONI data

Fig. 4. Average overlap ratios on NIREP and LONI datasets. Two groupwise registration algorithms (i.e., our method in red and the conventional method in blue) are performed on NIREP and LONI datasets, with the average overlap ratios of all ROIs (at different registration stages) reported in (a) and (b), respectively.

The calculation of overlap ratio is similar as those described above. Figs. 4(a) and 4(b) show the average overlap ratios achieved by the two methods on NIREP and LONI datasets, respectively. It is obvious that our method with sharp mean (in red curves) outperforms the conventional method (in blue curves) during the whole registration procedure.

4 Conclusion

We have demonstrated the importance of keeping the sharp group mean image during the groupwise registration procedure, in order to improve the overall registration performance of the whole population. Specifically, we generalize a popular unbiased groupwise registration method by presenting a new objective function to achieve the groupwise registration with the sharp mean. By using various datasets with manually labeled ROIs, we have shown that our method can perform better than the conventional groupwise registration method, in terms of consistently aligning the ROIs across different subjects. We also demonstrate that our method can eventually produce a much sharper group mean image than the conventional method.

References

1. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)
2. Fletcher, P.T., Venkatasubramanian, S., Joshi, S.: The geometric median on Riemannian manifolds with application to robust atlas estimation. NeuroImage 45, 143–152 (2009)
3. Hamm, J., Davatzikos, C., Verma, R.: Efficient large deformation registration via geodesics on a learned manifold of images. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 680–687. Springer, Heidelberg (2009)
4. Park, H., Bland, P.H., Hero, A.O., Meyer, C.R.: Hero. 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. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45, 61–72 (2009)

Lung Lobar Slippage Assessed with the Aid of Image Registration

Youbing Yin[1], Eric A. Hoffman[2,⋆], and Ching-Long Lin[1]

[1] Department of Mechanical & Industrial Engineering
[2] Departments of Biomedical Engineering, Radiology, and Medicine
The University of Iowa, Iowa City, IA 52242
{youbing-yin,eric-hoffman,ching-long-lin}@uiowa.edu

Abstract. We present a registration algorithm that can handle the discontinuity of deformation with an ultimate goal to investigate how pulmonary lobes deform to accommodate chest wall shape changes. We first show that discontinuities can exist in both normal and tangent directions. Such discontinuities are accounted for by a spatially varying diffusive regularization which restricts smoothing inside objects. Meanwhile, a distance term is combined with the sum of squared intensity differences (SSD) to explicitly match corresponding interfaces and intensity patterns. The capability of this new method is demonstrated using two-dimensional (2-D) synthetic examples with complete or incomplete "fissures" and three-dimensional (3-D) computed tomography (CT) lung datasets.

1 Introduction

Nonrigid image registration has become an important non-invasive tool to assess organ motion from a pair of medical images. Most earlier registration algorithms assume or imply a continuous and smooth deformation field. However, those algorithms might introduce unphysiological artifacts near object interfaces if the adjacent objects slip against each other. Thus, developing physiologically meaningful registration algorithms that account for discontinuities is becoming an important issue [1,2,3,4,5,6]. Of particular interest to this paper is discontinuity near lobar fissures in human lungs.

It is well known that human lungs are separated into five lobes by means of infolded reflections of the visceral pleura and lobes can slide against the chest wall and adjacent lobes [7]. Such motion may provide a means to reduce lung parenchymal distortion and avoid regions of high local stress. Conversely, it has previously been shown that lobar fissures fibrose and essentially disappear in animals (sloths) which undergo very little chest wall shape changes [8]. Recently, some registration algorithms have been proposed to account for discontinuity of deformation near lung borders or lobar fissures [2,3,4,5]. In [2,3,4], discontinuities are accounted for by registering regions of interest alone or separately with segmentation masks. Voxels outside of the region of interest are not taken into

⋆ E.A. Hoffman is shareholder in VIDA Diagnostics, Inc.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 578–585, 2010.

consideration during registration or are set to a uniform value before registration to form a high intensity contrast on the border. By this means these methods avoid inter-object regularization and match corresponding borders with an implicit penalty of the high intensity contrast between object and "background". However, such mask-based registration methods require a complete interface. Commonly, incomplete lobar fissures are observed in human lungs and the degree of incompleteness ranges from nearly complete absence to nearly complete presence of the fissure [9]. This means that lobes might be only partially separated and, thus, mask-based registration methods discussed above would be incapable of dealing with incomplete fissures. Different from mask-based registration, Kabus [1] introduced a spatially dependent weighting function to suppress the influence of smoothness near the interfaces of neighboring objects, thus, allowing for opening or closing of a gap between regions. Recently, Schmidt-Richberg *et al.* [5,6] proposed a direction-dependent regularization that restricts smoothing in normal directions while allowing discontinuity in tangent directions. An underlying assumption for their approach [5] is that normal components of displacements on the interface of neighboring objects are continuous. However, this assumption might not be true for lobar fissures since normal directions of fissures change during respiration, leading to discontinuities in both normal and tangent directions (it will be explained in details later). Furthermore, Schmidt-Richberg *et al.*'s approach cannot ensure corresponding interfaces to be matched between two images if the intensity contrast near the interface is weak, such as in the case for CT images of lobar fissures.

In this paper, we present a novel registration algorithm to deal with discontinuity of deformation occurring near lobar fissures. We show that discontinuities can exist in both normal and tangent directions and then present the new registration to account for such discontinuities. The proposed method is validated using two-dimensional (2-D) synthetic examples with complete or incomplete "fissures" and then applied to three-dimensional (3-D) computed tomography (CT) lung datasets.

2 Methods

2.1 Background

Given a pair of N-dimensional (N-D) images R and F, referred to as the reference and floating images, the goal of a registration algorithm is to find a deformation field to match images R and F. If there is only one homogeneous object, we should expect a smooth and continuous deformation field. However, slippage or discontinuity may occur on interfaces if there are more than one object. Fig. 1 illustrates three sliding cases, where the interfaces before and after motions are denoted by $\partial\Omega_1$ and $\partial\Omega_2$, respectively. A pair of adjacent points p_+ and p_- on $\partial\Omega_1$ is deformed into corresponding points q_+ and q_- on $\partial\Omega_2$. Due to slippage, q_+ and q_- are separated but still locate on $\partial\Omega_2$. In left panel of Fig. 1, interfaces before and after motion are planes with unchanged normal directions; in the middle both interfaces are planes but normal directions are different; and in the

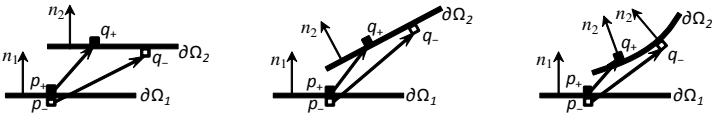


Fig. 1. A sketch illustrating sliding in three cases. Left: interfaces before and after motions are planes with unchanged normal directions; Middle: interfaces before and after motions are planes but with changed normal direction; Right: an initial plane interface is deformed into a curved surface after motion.

right an initial plane interface is deformed into a curved surface with different normal direction at q_+ and q_-. It is clear that normal components (with respect to n_1) of displacements for $p+$ and $p-$ are continuous only for the first case while discontinuities exist in both normal and tangent directions for the other two cases where the normal directions change. Based on these observations, we should not expect continuous normal components of displacement near lobar fissures since their motions are usually complex with changed normal directions. Thus, a new registration method is needed to investigate discontinuity of deformation between lung lobes.

2.2 Registration Method

Image registration is a process of determining an optimal deformation field to match images R and F. For non-parametric registration algorithms, deformation field is usually expressed in the form of displacement field $s : p \rightarrow s(p)$ for each voxel p. $s(\cdot)$ is usually solved by minimizing a global energy

$$E = Sim(R, F, s) + \frac{1}{\sigma_T^2} Reg(s), \tag{1}$$

where Sim is the similarity measure that defines how well images R and F match. Reg is the regularization term that avoids unstable and non-smooth solutions and σ_T is a weight coefficient to control the amount of regularization.

In order to capture discontinuities of deformation in both the normal and tangent directions, as we observe in Section 2.1, we need to suppress the regularization near the interface of neighboring objects while maintain regularization inside objects. Meanwhile, we have to ensure corresponding interfaces from images R and F to be matched to prevent gaps or intersections. For the first point, we adopt a spatially variable diffusive regularization:

$$Reg(s) = \sum_{l=1}^{N} \int_{\Omega} \beta(p) \|\nabla s_l(p)\|^2, \tag{2}$$

where $\beta(p)$ is a varying weight coefficient and s_l is the lth component of s. A desirable β should restrict the regularization inside objects and one of choices is a Dirac-shaped function [1,5]: $\beta(p) = 1/(1 + \alpha n \exp(-\alpha \phi_R^2(p))$, where α and n

are control parameters and we choose $\alpha = n = 1000$ in this work. $\phi_R(p)$ is the Eucilidean distance at voxel p to its closest point on the interface in image R.

For the similarity term in Eq. (1), we combine the sum of squared intensity differences (SSD) and a distance penalty term. The distance term measures distances of all points on an interface in image R to the corresponding interface in image F and the idea is similar to surface matching algorithms using distance transformation [10]. The combined similarity measure is in the form of

$$Sim(R, F, s) = \int_{\Omega} \|R - F \circ s(p)\|^2 + \gamma \int_{\partial\Omega} \|\phi_F \circ s(p)\|^2, \tag{3}$$

where ϕ_F is the Eucilidean distance field from the corresponding interface in image F and γ is a weight coefficient.

Although Eq. (1) could be solved directly [1], the coupling of the similarity and regularization terms usually makes it computationally complicated. In this work, we solve Eq. (1) based on a Demons framework [11] and rewrite it as

$$E = \int_{\Omega} \|R - F \circ c\|^2 + \gamma \int_{\partial\Omega} \|\phi_F \circ c\|^2 + \frac{1}{\sigma_x^2} \int_{\Omega} \|c - s\|^2 + \frac{1}{\sigma_T^2} \sum_{l=1}^{N} \int_{\Omega} \beta \|\nabla s_l\|^2, \tag{4}$$

where the third term is called as the correspondence term and σ_x accounts for a spatial uncertainty. c is usually expressed in a composition of s and a small updated deformation u: for example, $s \circ (\mathrm{Id} + u)$. By introducing the correspondence term, the coupled optimization can be decoupled into a two-step procedure. It starts from an initial displacement field s^0. The first step is to solve u^t by minimizing the first three terms with s^t being given. The second step is to solve s^{t+1} by minimizing the last two terms with $c^t = s^t \circ (\mathrm{Id} + u^t)$ being given.

3 Results

3.1 2-D Synthetic Experiments

The performance of the proposed algorithm is first evaluated by 2-D synthetic images. Three synthetic cases are designed with complete or incomplete internal interfaces, denoting fissures, as shown in Fig. 2. A line A-A is chosen for each case for comparison between registration-predicted displacements and exact solutions.

The deformation field and intensity patterns are designed to deform "fissures" into designated positions while keep undeformed outside of objects. Fig. 3 shows intensity patterns (gray-scale images with intensity range [0, 255], 128×128 pixel size, and $1 \times 1\,\mathrm{mm}^2$ pixel spacing), deformation fields, and registration results for the three cases above. The reference images are presented on the first three panels in the 1st row, respectively, and they are all registered to the same floating image as shown on the last panel in the 1st row. The warped image, difference image and deformation fields are shown from the 2nd to 4th rows, respectively, with each column corresponding to each case in Fig. 2. Comparisons of $x-$ and y-components of displacements (u, v) between predicted results and exact solutions along A-A are shown from left to right in Fig. 4 for the three cases,

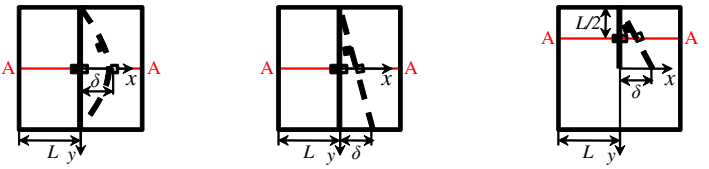


Fig. 2. Sketches for 2-D synthetic cases. Left: Complete Plane-Parabolic, a complete plane "fissure" is deformed into a parabolic surface; Middle: Complete Plane-Plane, a complete plane "fissure" is deformed with changed orientation; Right: Incomplete Plane-Plane, an incomplete plane "fissure" is deformed with changed orientation. The reference line A-A is marked in red.

respectively. It can be seen that overall the proposed approach could recover the true deformation field for all the three cases. Discontinuities of both normal (u) and tangent (v) components of displacements for cases with complete or incomplete "fissures" are captured and are consistent with the ground truth.

3.2 3-D CT Experiments

In this section, we apply the proposed algorithm to 3-D CT lung datasets with complete lobar fissures. Three pairs of datasets from three human subjects (2 males, 1 female, age: 20-26 years, normal non-smokers) are used. All datasets were acquired with a Siemens Sensation 64 multi-detector row CT scanner (Forchheim, Germany) during breath-holding at 60% and 80% of vital capacity under an approved protocol by our institutional review board. Subjects were studied as part of a project seeking to establish a normative lung atlas. We chose these two lung volumes somewhat arbitrarily from datasets with multiple volumes. The volume change (about 0.8 liters) was only slightly greater than a typical tidal volume and thus was suitable for this early evaluation. In-plane dimension is 512×512 with an approximate spatial resolution of $0.6 \times 0.6\,\text{mm}^2$. The z-direction dimension ranges from 600 to 800 with a spatial spacing of 0.5 mm. Complete lobar fissures are detected by using the Pulmonary Workstation 2 (VIDA Diagnostics, Iowa, United States) with the algorithm derived in [12]. Similar to [3], we limit our current analysis to the upper and lower lobes of the left lung. Original dataset are cropped to include left lung only. Although we notice that changes in voxel intensity due to respiration might affect registration accuracy [3,13,14], we expect that the influence is small here since volume differences are relatively small. Landmarks located at vessel bifurcations are used to assess registration accuracy. Each registration pair has 40–60 landmarks and they are distributed in the whole left lung. The mismatch errors are 1.25 ± 1.67 (6.49 ± 5.29), 0.80 ± 0.39 (5.80 ± 3.71), and 1.70 ± 1.35 (8.31 ± 2.82) for three registration pairs, respectively, where units are mm and numbers in $(\cdot)$ denote errors before registrations. Fig. 5 shows the floating image, reference image, warped image and displacement field from left to right from one subject, representative of the three subjects studied. Discontinuity near the fissure is

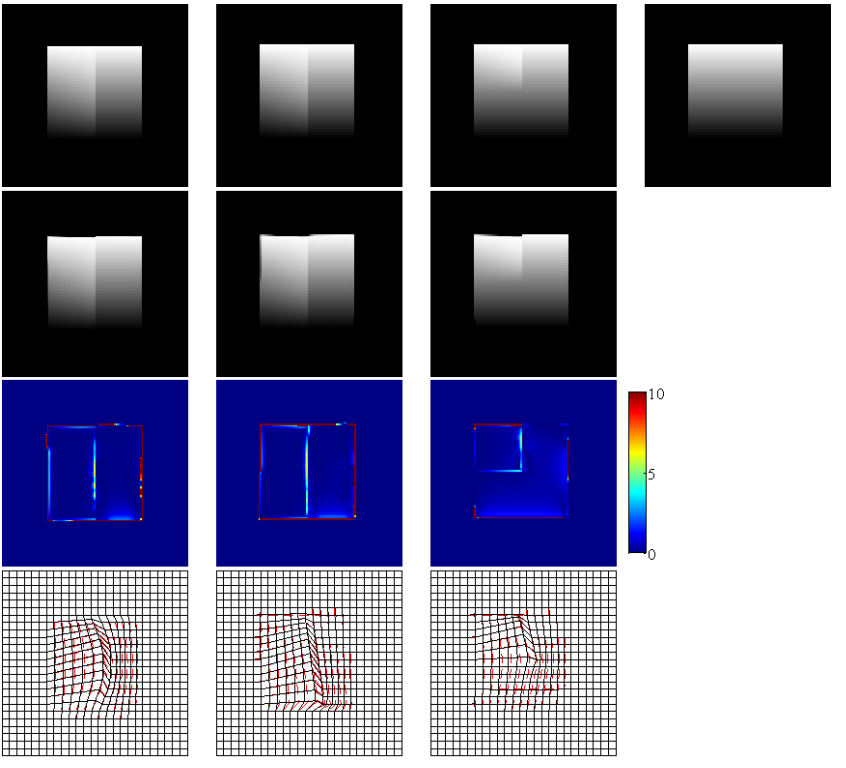


Fig. 3. Results for cases (from 1st to 3rd column): "Complete Plane-Parabolic", "Complete Plane-Plane", "Incomplete Plane-Plane". Reference images are shown and are registered to the same floating image (the last one) in the 1st row. The warped image, difference image, and overlay of exact (red dashed) and predicted (black solid) deformation grids are presented from the 2nd to 4th rows.

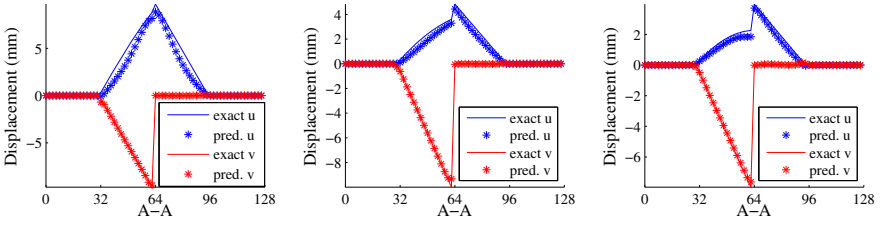


Fig. 4. Comparisons of u and v-components of displacements along A-A for three cases: left, "Complete Plane-Parabolic"; middle, "Complete Plane-Plane"; right, "Incomplete Plane-Plane"

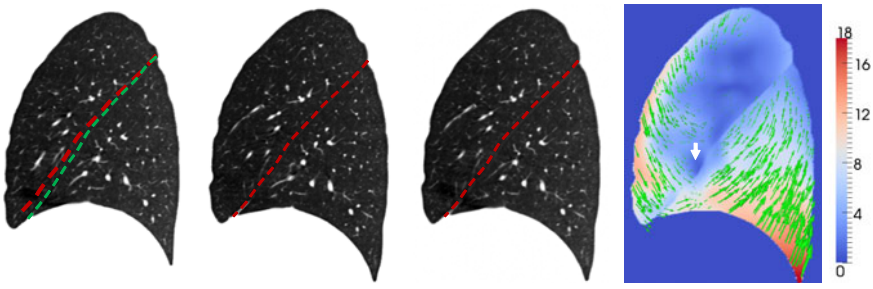


Fig. 5. 3-D lung CT experiment. From left to right: the floating image, reference image, warped image, and displacement field. Oblique fissure is highlighted in green for the floating image and in red for the reference and warped images. The white arrow in the right panel shows a region of low displacement.

clearly visualized. The white arrow in the right panel shows a region with low displacement where motion is restricted by the proximity of the heart.

4 Discussion and Summary

In this work, we proposed a novel registration method to capture discontinuities of deformation. We first show that discontinuities can exist in both normal and tangent directions. Such discontinuities are accounted for by a spatially variable diffusive regularization. In addition, we incorporate a distance penalty term into SSD to explicitly match both intensity patterns and interfaces. The new method is evaluated using 2-D synthetic examples with complete or incomplete "fissures", and results show that it is capable of capturing discontinuity of deformation in both normal and tangent directions. The method is then applied to 3-D CT lung datasets with complete fissures. An important next step is to apply the proposed method to datasets with both complete and incomplete fissures to investigate the influence of degree of fissure incompleteness on regional lung mechanics, which will help develop a breathing lung model for computational fluid dynamics simulation of pulmonary air flow [15,16,17].

In the current work, fissure interfaces for both images to be matched are required and a penalty term with the Eucilidean distance is used to match the corresponding interfaces. The interfaces are assumed accurate and investigation of potential effects of their accuracy on registration results would be necessary for future work. An additional issue related to the penalty term is how to prevent gaps or intersections during registrations. Currently we have not checked such uplausible results but we did checked distances of all voxels on the interfaces to ensure that most of them are within one voxel by changing the weight coefficient.

Acknowledgments

This work was supported in part by NIH Grants R01-HL-064368, R01-EB005823, S10-RR-022421, and R01-HL-094315.

References

1. Kabus, S.: Multiple-material variational image registration. PhD thesis, University of Lübeck (2007)
2. Wu, Z., Rietzel, E., Boldea, V., Sarrut, D., Sharp, G.C.: Evaluation of deformable registration of patient lung 4DCT with subanatomical region segmentations. Med. Phys. 35(2), 775–781 (2008)
3. Ding, K., Yin, Y., Cao, K., Christensen, G.E., Lin, C.-L., Hoffman, E.A., Reinhardt, J.M.: Evaluation of lobar biomechanics during respiration using image registration. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 739–746. Springer, Heidelberg (2009)
4. Xie, Y., Chao, M., Xing, L.: Tissue feature-based and segmented deformable image registration for improved modeling of shear movement of lungs. Int. J. Radiation Oncology Biol. Phys. 74(4), 1256–1265 (2009)
5. Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Handels, H.: Slipping objects in image registration: improved motion field estimation with direction-dependent regularization. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 755–762. Springer, Heidelberg (2009)
6. Schmidt-Richberg, A., Ehrhardt, J., Werner, R., Handels, H.: Direction-dependent regularization for improved estimation of liver and lung motion in 4D image data. In: SPIE Medical Imaging, vol. 7623, pp. 76232Y (2010)
7. Rodarte, J.R., Hubmayr, R.D., Stamenovic, D., Walters, B.J.: Regional lung strain in dogs during deflation from total lung capacity. J. Appl. Physiol. 58, 164–172 (1985)
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. Lubner, M.G.: The incomplete fissure sign. Radiology 247(2), 589–590 (2008)
10. Betke, M., Hong, H., Thomas, D., Prince, C., Ko, J.P.: Landmark detection in the chest and registration of lung surfaces with an application to nodule registration. Med. Image Anal. 7, 265–281 (2003)
11. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45, S61–S72 (2009)
12. Ukil, S., Reinhardt, J.: Anatomy-guided lung lobe segmentation in x-ray ct images. IEEE Trans. Med. Imaging 28(2), 202–214 (2009)
13. Yin, Y., Hoffman, E.A., Lin, C.-L.: Mass preserving nonrigid registration of CT lung images using cubic B-spline. Med. Phys. 36(9), 4213–4222 (2009)
14. Yin, Y., Hoffman, E.A., Lin, C.-L.: Local tissue-weight-based nonrigid registration of lung images with application to regional ventilation. In: SPIE Medical Imaging, vol. 7262, pp. 72620C (2009)
15. Lin, C.-L., Tawhai, M.H., McLennan, G., Hoffman, E.A.: Characteristics of the turbulent laryngeal jet and its effect onairflow in the human intra-thoracic airways. Respir. Physiol. Neurobiol. 157(2-3), 295–309 (2007)
16. Lin, C.-L., Tawhai, M.H., McLennan, G., Hoffman, E.A.: Multiscale simulation of gas flow in subject-specific models of the human lung. IEEE Eng. Med. Biol. Mag. 28(3), 25–33 (2009)
17. Yin, Y., Choi, J., Hoffman, E.A., Tawhai, M.H., Lin, C.-L.: Simulation of pulmonary air flow with a subject-specific boundary condition. J. Biomech. (2010), doi: 10.1016/j.jbiomech.2010.03.048

Generalization of Deformable Registration in Riemannian Sobolev Spaces

Darko Zikic[1], Maximilian Baust[1], Ali Kamen[2], and Nassir Navab[1]

[1] Computer Aided Medical Procedures (CAMP), TU München, Germany
[2] Siemens Corporate Research (SCR), Princeton, NJ, USA

Abstract. In this work we discuss the generalized treatment of the deformable registration problem in Sobolev spaces. We extend previous approaches in two points: 1) by employing a general energy model which includes a regularization term, and 2) by changing the notion of distance in the Sobolev space by problem-dependent Riemannian metrics. The actual choice of the metric is such that it has a preconditioning effect on the problem, it is applicable to arbitrary similarity measures, and features a simple implementation. The experiments demonstrate an improvement in convergence and runtime by several orders of magnitude in comparison to semi-implicit gradient flows in L^2. This translates to increased accuracy in practical scenarios. Furthermore, the proposed generalization establishes a theoretical link between gradient flow in Sobolev spaces and elastic registration methods.

Keywords: Deformable Registration, Sobolev Spaces, Riemannian Manifolds, Preconditioning.

1 Introduction

The goal of intensity-based deformable registration is the minimization of the similarity measure between a source image I_S, warped by a deformation ϕ, and a target image I_T. The d-dimensional images are defined as $I : \Omega \subset \mathbb{R}^d \to \mathbb{R}$, and the deformation $\phi \in \mathcal{H}$, $\phi : \Omega \to \mathbb{R}^d$ is defined in a Hilbert space $\mathcal{H}$, and is expressed in terms of the identity transformation and the displacement $u \in \mathcal{H}$ as $\phi = \mathrm{Id} + u$. The problem is generally modeled as a minimization of an energy consisting of a similarity measure E_D and a regularization term E_R, that is

$$E(\phi) = E_D(\phi) + \lambda E_R(\phi) \ . \tag{1}$$

There are numerous choices for E_D and E_R, cf. [1,2]. The optimization of (1) by a gradient flow in $\mathcal{H}$ consists of iterative application of the evolution rule

$$\partial_t \phi = -\tau \cdot \nabla_{\mathcal{H}} E(\phi) \qquad \text{// update as negative multiple of gradient} \tag{2}$$

$$\phi = \phi \oplus \partial_t \phi \qquad \text{// application of the update} \tag{3}$$

Here, $\nabla_{\mathcal{H}}$ is the gradient in $\mathcal{H}$, and $\oplus$ defines the appropriate update operation [3].

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 586–593, 2010.

A common choice for $\mathcal{H}$ is the space of square integrable functions L^2, cf. e.g. [2]. In [4,3], Sobolev spaces are discussed as a choice for $\mathcal{H}$. Compared to L^2, Sobolev spaces have the advantage to contain only functions with certain regularity properties, which is favorable for deformable registration. Due to this inherent regularity, problems which are ill-posed in L^2 can be well-posed in Sobolev spaces. This is the case for the minimization of (1) with $\lambda = 0$. In L^2, this is an ill-posed problem and a regularization term E_R is necessary to allow a numerical treatment, while the problem is well-posed in an appropriate Sobolev space. This has been recognized and put to use in [4,3]. In these approaches, no regularization term E_R is used, and the required smoothness is instead achieved by choosing an appropriate "geometric setting" by employing Sobolev spaces [3].

The motivation behind omitting E_R in these works was to enable the recovery of large deformations, which can be prohibited by strong regularization. This motivation is the same as in the so-called fluid approaches [5], which were shown to be equivalent to minimization of E_D in a suitable Sobolev space [4]. The increased flexibility of fluid approaches comes at the cost of inhibited propagation of ϕ into homogeneous and low contrast image regions, which is otherwise achieved by E_R in the general model (1). Thus, depending on input data, omitting E_R altogether can present a drawback.

In this work, we generalize previous approaches for deformable registration in Sobolev spaces in two points. The first point is that we consider the complete energy from (1), including the regularization term E_R. Compared to traditional fluid-type Sobolev-based approaches (obtained for $\lambda = 0$), the inclusion of E_R is of advantage for treatment of images with large low contrast regions. Besides, this generalization provides a theoretically interesting interpretation of elastic registration methods, by identifying the semi-implicit time discretization of a gradient flow in L^2 (cf. e.g. [2]) as steepest descent in a suitable Sobolev space. In [6], regularization is used together with a Sobolev space setting. This is conceptually quite a different approach and the regularization is performed on velocities, adding a temporal dimension to the problem.

The second generalization is the use of problem-dependent Riemannian metrics for definition of Sobolev spaces. The use of Sobolev spaces in [4,3] is directed at restricting the space of deformations to a certain class, such as diffeomorphisms. Our approach builds on these results by preserving the geometric setting, and extends it by changing the notion of distance in these spaces by Riemannian metrics. This provides us with a flexible theoretical framework, allowing to change the properties of the underlying space, such that its numerical treatment becomes more efficient. We design the Riemannian metrics based on the specific problem and the input data, such that the metric has a preconditioning effect on the optimization problem. We present a strategy to generate the Riemannian metric for arbitrary similarity measures used for registration of medical images, and show that the resulting algorithm exhibits a significantly improved convergence, resulting in much shorter overall runtimes.

By the rationale for the metric choice, our method relates to work on preconditioned gradient descent [7,8]. In contrast to these methods which assume a

mono-modal scenario and employ SSD, the proposed technique makes no such assumptions, and is shown to work in multi-modal settings with MI.

It is important to note that while the choice of $\mathcal{H}$ influences the path of the optimization process and the resulting local minimum, it does not affect the definition of (1), such that the optimization operates on the same energy. Thus, by selection of an appropriate metric, we can hope to construct "shorter" paths on the given energy, resulting in more efficient methods.

2 Method

After a brief introduction of Sobolev spaces, and a discussion of previous uses for deformable registration in Sec. 2.1, we generalize the standard approach by using Riemannian metrics for definition of Sobolev spaces in Sec. 2.2. Section 2.3 motivates the selection of metrics based on preconditioning, and in Sec. 2.4, we propose the construction of such metrics for registration purposes. Section 2.5 highlights relations of some well-known methods to the proposed approach.

2.1 Sobolev Spaces and Sobolev Gradients

The following discussion is based on [9,10]. The Sobolev space H^k on $\mathbb{R}^n$ is a Hilbert space of functions whose derivatives up to the order k are square integrable, that is $H^k = \{f : \|f\|_{H^k} < \infty\}$, with

$$\|f\|_{H^k} = \left(\sum_{i=0}^{k} \langle f^{(i)}, f^{(i)} \rangle_{L^2} \right)^{\frac{1}{2}} \equiv \langle f, f \rangle_{H^k}^{\frac{1}{2}} \ . \tag{4}$$

The H^k scalar product can be written in terms of the L^2 scalar product with the use of the vector-valued differential operator $\mathcal{L} : \mathcal{H} \to \mathcal{H}^k$, $\mathcal{L} = (D^0 \dots D^k)$, consisting of differential operators D^i of order i, as

$$\langle f, f \rangle_{H^k} = \sum_{i=0}^{k} \langle f^{(i)}, f^{(i)} \rangle_{L^2} = \langle \mathcal{L}f, \mathcal{L}f \rangle_{L^2} = \langle \mathcal{L}^*\mathcal{L}f, f \rangle_{L^2} \ , \tag{5}$$

where $\mathcal{L}^*$ is the adjoint of $\mathcal{L}$, and the differential operator $\mathcal{L}^*\mathcal{L}$ has the form

$$\mathcal{L}^*\mathcal{L} = \sum_{i=0}^{k} (-1)^i \Delta^i \ . \tag{6}$$

To express the Sobolev gradient by the L^2 gradient, we use the definition of gradient in the space $\mathcal{H}$ as the entity $\nabla_{\mathcal{H}} f$ which can be used to represent the directional derivative $\partial_h f$ by $\partial_h f = \langle \nabla_{\mathcal{H}} f, h \rangle_{\mathcal{H}}$. By applying this for $\mathcal{H} = L^2$ and $\mathcal{H} = H^k$, we can equate the resulting right hand sides, and use (5), yielding

$$\langle \nabla_{L^2} f, h \rangle_{L^2} = \langle \nabla_{H^k} f, h \rangle_{H^k} = \langle \mathcal{L}^*\mathcal{L} \nabla_{L^2} f, h \rangle_{L^2} \ . \tag{7}$$

From this, we can express the Sobolev gradient in terms of the L^2 gradient as

$$\nabla_{H^k} f = (\mathcal{L}^*\mathcal{L})^{-1} \nabla_{L^2} f \ . \tag{8}$$

In previous works, some modifications of the operator $\mathcal{L}^*\mathcal{L}$ have been proposed. In [3], a general form

$$\mathcal{L}_\alpha^*\mathcal{L}_\alpha = (\alpha\mathcal{L})^*(\alpha\mathcal{L}) = \sum_{i=0}^{k}(-1)^i \alpha_i \Delta^i \ , \tag{9}$$

is considered, with $\alpha_i \in \mathbb{R}$. Two specific instances of (9) are considered in more detail in [3]. We focus on the first one with $\alpha_0 = 1$, and $\alpha_1 = \alpha \in \mathbb{R}$, resulting in

$$\mathcal{L}_\alpha^*\mathcal{L}_\alpha = \mathrm{Id} - \alpha\Delta \ . \tag{10}$$

2.2 Generalization of Sobolev Gradients to Riemannian Manifolds

We generalize (4), by introducing Riemannian metric tensors M_i to the single scalar products by

$$\|f\|_{H_M^k} = \left(\sum_{i=0}^{k} \langle M_i f^{(i)}, f^{(i)} \rangle_{L^2}\right)^{\frac{1}{2}} \ , \tag{11}$$

thus treating the single derivatives in Riemannian manifolds. In contrast to the use of scalars α_i in (9), we employ Riemannian metric tensors $M_i = M_i'^* M_i'$, which are by definition symmetric positive definite, and vary smoothly in the space of deformations. With the operator $\mathcal{L}_M = (M_0' D^0 \dots M_k' D^k)$, we get

$$\mathcal{L}_M^*\mathcal{L}_M = \sum_{i=0}^{k}(-1)^i \ {\nabla^i}^\top M_i \nabla^i \ . \tag{12}$$

According to (8), the gradient with respect to H_k^1 reads

$$\nabla_{H_M^k} f = \left(\mathcal{L}_M^*\mathcal{L}_M\right)^{-1} \nabla_{L^2} f \ . \tag{13}$$

Please note that we do not change the class of functions contained in the Sobolev space, but only the notion of distance, because the positive definiteness of all M_i ensures the existence of $0 < c, C < \infty$ such that

$$c \cdot \|f\|_{H^k} \leq \|f\|_{H_M^k} = \left(\langle \mathcal{L}_M f, \mathcal{L}_M f \rangle_{L^2}\right)^{\frac{1}{2}} \leq C \cdot \|f\|_{H^k} \ . \tag{14}$$

Thus $H^k = \{f : \|f\|_{H^k} < \infty\} = \{f : \|f\|_{H_M^k} < \infty\} = H_M^k$.

In the remainder of the paper, we will restrict our treatment to $k = 1$. The obtained results are however readily transferable to general settings. Corresponding to (10), for the generalized formulation of H_M^1 we get

$$\mathcal{L}_M^*\mathcal{L}_M = M_0 - \mathrm{div}(M_1 \nabla) \ . \tag{15}$$

In summary, the computation of the update in (2) for H_M^1 is now based on

$$\nabla_{H_M^1} E(\phi) = \left(M_0 - \mathrm{div}(M_1\nabla)\right)^{-1} \left(\nabla_{L^2} E_D(\phi) + \lambda \nabla_{L^2} E_R(\phi)\right) \ . \tag{16}$$

2.3 Selection of Metric Based on Preconditioning

The general formulation from (16) provides us with a framework in which we can choose the metrics M_i such that the resulting algorithms have advantageous properties. We propose to select these metrics such that the convergence of the algorithm is improved. This can be done by interpreting (13) as the result of preconditioning of the given problem in the L^2 setting. Preconditioning is a standard technique for improvement of convergence rate [11]. To this end, consider the second-order Taylor approximation of f in a Hilbert space $\mathcal{H}$

$$f(x+\alpha h) = f(x) + \alpha\langle h, \nabla_{\mathcal{H}} f(x)\rangle_{\mathcal{H}} + \frac{\alpha^2}{2}\langle h, H_{\mathcal{H}}(f)h\rangle_{\mathcal{H}} + \mathcal{O}(\alpha^3) \ . \tag{17}$$

For a critical point x' with $\nabla f(x') = 0$, the first order term in (17) disappears and H dominantly describes the shape of f about x', so that the condition of H has a direct impact on the convergence of gradient-based methods, see also [10].

As for the gradient in Eq. (8), we can express $H_{H^1_M}$ in terms of H_{L^2} by

$$H_{H^1_M} = \left(\mathcal{L}_M^* \mathcal{L}_M\right)^{-1} H_{L^2} \ . \tag{18}$$

Now, we can influence the condition of $H_{H^1_M}$ by an appropriate choice of $\mathcal{L}_M^*\mathcal{L}_M$ as an approximation to H_{L^2}, thus improving the convergence properties. The choice of $\mathcal{L}_M^*\mathcal{L}_M$ should be simple, efficient, and numerically stable.

2.4 Metric Selection for Deformable Registration

While the structure of $E_D = \int \mathcal{D}_D$ is general as the integration can be performed over the spatial domain (e.g. for SSD), or the intensity domain for statistical measures (CC, CR, MI), the regularization is mostly formulated as a least-squares term in the spatial domain, so that with a differential operator $\mathcal{D}_R$ we can rewrite (1) as $E(\phi) = E_D(\phi) + 1/2 \cdot \lambda\langle \mathcal{D}_R u, \mathcal{D}_R u\rangle_{\mathcal{H}}$. The corresponding L^2 Hessian reads

$$H_{L^2}(E) = H_{L^2}(E_D) + \lambda H_{L^2}(E_R) = H_{L^2}(E_D) + \lambda \mathcal{D}_R^* \mathcal{D}_R \ . \tag{19}$$

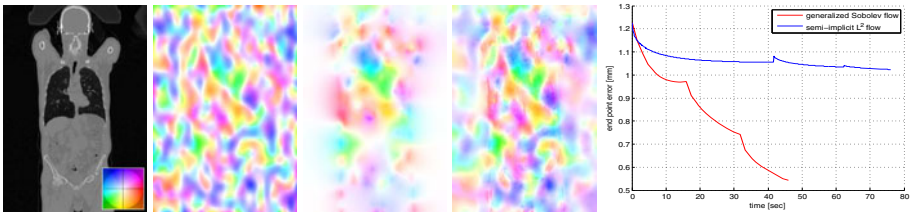
2.4.1 Preconditioning of Regularized Energies by Sobolev Gradients

The first observation is that the use of Sobolev spaces can result in preconditioning of the general energy from (1). To this end, the Sobolev space must be based on the same differential operator $\mathcal{D}_R$ which is employed for the regularization, that is $\mathcal{L} = (\mathrm{Id}, \mathcal{D}_R)$. Thus, with $M_0 = \mathrm{Id}$ and $M_1 = \lambda \mathrm{Id}$ in (15) we get

$$\mathcal{L}_M^* \mathcal{L}_M = \mathrm{Id} + \lambda \mathcal{D}_R^* \mathcal{D}_R \ , \tag{20}$$

which provides a preconditioner for general energies with regularization terms, since $\mathcal{L}^*\mathcal{L}$ from (20) is an approximation to $H_{L^2}(E)$ in (19). This tells us that the steepest descent can be expected to converge much faster in Sobolev spaces than in L^2, since it can be seen as a preconditioned version.

Please note that Sobolev spaces based on (20) are of the form (9) and were employed in [4,3]. In these works however, the above preconditioning argument does not hold since no regularization term E_R is employed.



(a) source (b) GT exampl. (c) L^2 displ. (d) our result (e) convergence plot

Fig. 1. Mono-modal random study. Results in (e) are the mean of 100 trials, w.r.t. computation time. Displacements in (b)-(d) are color-coded, c.f. (a). The proposed method clearly outperforms the semi-implicit L^2 flow in terms in speed and accuracy.

2.4.2 Further Preconditioning by Generalized Sobolev Gradients

The next step to improve the condition of $H_{H^1_M}$ in (18) is to choose M_0 such that $M_0 \approx H_{E_D}$. It is crucial that: 1) the approximation can be computed efficiently, and 2) the approximation is applicable to arbitrary similarity measures. Simple standard preconditioning techniques such as Jacobi preconditioning ($M_0 = \mathrm{diag}(H)$) proved ineffective in our experiments.

We propose to compose the Riemannian metric M_0 as a block diagonal matrix, where each $d \times d$ block $M_0(x) \in \mathbb{R}^d$ corresponds to a spatial position $x \in \Omega$. The single blocks $M_0(x)$ are scaled structure tensors of the energy term for single spatial positions $x \in \Omega$, evaluated at the current estimate u, that is

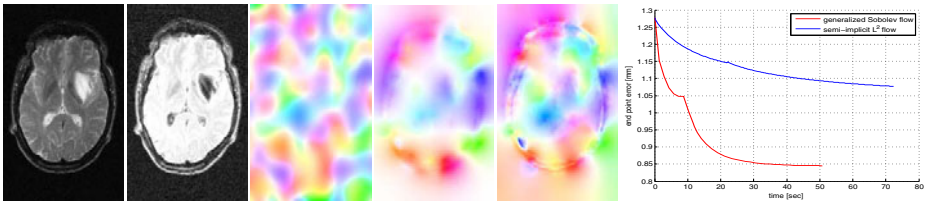
$$M_0(x) = \sigma \mathrm{Id} + \frac{1}{\|\nabla_{L^2} E_D(u)(x)\|} \nabla_{L^2} E_D(u)(x) \; \nabla_{L^2} E_D(u)(x)^\top \quad , \qquad (21)$$

where $\sigma \in \mathbb{R}$ is a small stabilization parameter, which assures the positiveness of M_0, and $\nabla_{L^2} E_D(u)(x)\mathbb{R}^d$ are the sub-vector of the energy gradient at single spatial positions $x \in \Omega$. This choice is applicable for arbitrary similarity measures, it results in a highly sparse, symmetric positive definite d-diagonal metric M_0, which is extremely efficient to compute as $\nabla_{L^2} E_D(u)$ is already calculated in every iteration. Furthermore, the solution of the resulting linear system with the operator $(M_0 - \lambda\Delta)$ can be performed by standard fast linear system solvers [12].

The motivation for the choice of (21) is based on the analysis of the Gauß-Newton optimization for an energy based on SSD and diffusion regularization. It can be shown that in this case, the effect of the preconditioning by the Gauß-Newton method can be seen as an approximate normalization of the magnitudes of the point-wise sub-vectors $\nabla_{L^2} E_D(u)(x)$. This choice is closely related to the analyses of the demons method provided in [13] and [14]. For more details please see the supplementary material at `http://campar.in.tum.de/personal/zikic/miccai2010/index.html`.

2.5 Relation to Other Methods

It is interesting to observe that some well-known methods can be seen as special cases of the proposed approach. Starting from the evolution rule derived in Eq.



(a) source (b) target (c) GT (d) L^2 displ. (e) our result (f) convergence plot

Fig. 2. Random study in a multi-modal setting, using MI, demonstrates applicability to statistical similarity measures. (b)-(e) depict entities from one trial.

(16), we note that for $M_0 = \mathrm{Id}$, and $M_1 = \lambda\tau\mathrm{Id}$ the Sobolev flow is equivalent to semi-implicit discretization of time in the Euler-Lagrange term arising in the L^2-based gradient flow approach [2]. Furthermore, the classical optical flow method by Horn and Schunck [15] can be seen as a generalized Sobolev flow, and is obtained by employing diffusion regularization, and SSD as similarity measure, with $M_0 = J_f^\top J_f$, with J_f being the Jacobian of $f = I_T - I_S \circ \phi$, and $M_1 = \lambda\mathrm{Id}$.

3 Evaluation

We perform 2D random studies in a controlled environment with known ground truth to demonstrate the improvement in convergence and precision, which result from the proposed approach. We compare the proposed method to the standard semi-implicit approach as described in Sec. 2.5. Per study, we perform 100 trials in each of which the source image is warped by a random ground truth deformation ϕ_{GT}, generated by B-Spline FFDs, with maximal displacements of 5mm. Method parameters (α, τ) for the semi-implicit approach are carefully tuned for best possible performance. We monitor the mean euclidean distance between ϕ_{GT} and the estimated deformation (end-point error) in every iteration. A standard multi-level scheme is employed.

The first study is performed on a CT image with SSD as similarity measure, cf. Fig. 1. We demonstrate the applicability of the proposed approach to statistical similarity measures by a study with MI in a multi-modal scenario (Fig. 2). To this end, we employ an MR-T2 image (from http://www.insight-journal.org/RIRE/), with intensities rescaled to $[0, 1]$ as I_S, and register it to $I_T = \widetilde{I_S} \circ \phi_{\mathrm{GT}}$ which includes a non-linear modification of intensities by $\widetilde{I_S}(x) = I_S(x) \cdot (1 - I_S(x))$, in order to simulate a multi-modal scenario. We observe a clear improvement in terms of convergence speed and the actual runtime for the proposed method. The effectively resulting accuracy is also drastically improved, especially in low gradient regions. This is consistent with our choice of the metric in 2.4.2. We observed the same behavior in experiments for SAD and CC as similarity measures. While the single iterations of the proposed method take longer than for the semi-implicit approach, due to the extreme improvement of convergence rate, far less iterations are needed, which results in a significant reduction of the overall

runtime. For example, the results in Fig. 1 feature 30 iterations for the proposed, and 550 for the semi-implicit method.

It is important to note that the decrease of the energy is very similar for both approaches (cf. supplementary material) . Based on the inspection of energy logs alone, the semi-implicit method might be considered converged even if the actual error is still significant, c.f. Figs. 1, 2. This premature convergence is a serious pitfall for real applications in which the actual error cannot be measured.

4 Summary

We propose a generalization of previous work on deformable registration in Sobolev spaces by using an explicit regularization term in the energy model, and by modification of the notion of distance by introduction of Riemannian metrics. The general framework in combination with the choice of the Riemannian metric based on the idea of preconditioning leads to a simple and yet powerful method, which outperforms flow strategies in L^2 in terms of speed, and improves the resulting accuracy, especially in low-gradient areas, thus preventing possible premature convergence in real applications.

References

1. Hermosillo, G., Chefd'Hotel, C., Faugeras, O.: Variational methods for multimodal image matching. International Journal of Computer Vision, IJCV (2002)
2. Modersitzki, J.: Numerical methods for image registration. Oxford Univ. Pr., Oxford (2004)
3. Chefd'hotel, C.: Geometric Methods in Computer Vision and Image Processing: Contributions and Applications. PhD thesis, ENS Cachan (2005)
4. Trouvé, A.: Diffeomorphisms groups and pattern matching in image analysis. International Journal of Computer Vision, IJCV (1998)
5. Christensen, G., Rabbitt, R., Miller, M.: Deformable templates using large deformation kinematics. IEEE Transactions on Image Processing, TIP (1996)
6. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision, IJCV (2005)
7. Haber, E., Modersitzki, J.: A multilevel method for image registration. SIAM Journal on Scientific Computing (2006)
8. Klein, S.: Optimisation methods for medical image registration. PhD thesis, Image Sciences Institute, UMC Utrecht (2008)
9. Neuberger, J.: Sobolev gradients and differential equations. Springer, Berlin (1997)
10. Renka, R.: A simple explanation of the sobolev gradient method (2006)
11. Nocedal, J., Wright, S.: Numerical optimization. Springer, Heidelberg (2000)
12. Saad, Y.: Iterative methods for sparse linear systems. SIAM, Philadelphia (2003)
13. Pennec, X., Cachier, P., Ayache, N.: Understanding the demon's algorithm: 3d non-rigid registration by gradient descent. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 597–605. Springer, Heidelberg (1999)
14. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage (2009)
15. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence (1981)

An Efficient EM-ICP Algorithm for Symmetric Consistent Non-linear Registration of Point Sets

Benoît Combès and Sylvain Prima

INSERM, U746, F-35042 Rennes, France
INRIA, VisAGeS Project-Team, F-35042 Rennes, France
University of Rennes I, CNRS, UMR 6074, IRISA, F-35042 Rennes, France
{bcombes,sprima}@irisa.fr
http://www.irisa.fr/visages

Abstract. In this paper, we present a new algorithm for non-linear registration of point sets. We estimate both forward and backward deformations fields best superposing the two point sets of interest and we make sure that they are consistent with each other by designing a symmetric cost function where they are coupled. Regularisation terms are included in this cost function to enforce deformation smoothness. Then we present a two-step iterative algorithm to optimise this cost function, where the two fields and the fuzzy matches between the two sets are estimated in turn. Building regularisers using the RKHS theory allows to obtain fast and efficient closed-form solutions for the optimal fields. The resulting algorithm is efficient and can deal with large point sets.

1 Introduction

The most popular methods in the literature for the non-rigid registration of two point sets are probably those extending the original ICP algorithm of Besl & McKay [1], such as the EM-ICP of Granger & Pennec [2] or the TPS-RPM of Chui & Rangarajan [3]. The common viewpoint of these methods is to consider the points of the first set as the means of a Gaussian Mixture Model (GMM) and the points of the other set as samples of this GMM. The unknown non-rigid transformation best superposing the two point sets can then be estimated according to the maximum likelihood principle and using (typically) the EM algorithm. This optimisation then boils down to a simple iterative estimation of fuzzy point-to-point correspondences (encoded in what is often termed the *match matrix*) and of the non-rigid transformation in turn. This simple two-step (point matching and transformation estimation) algorithm is very attractive, but inherently asymmetric, which makes it difficult to obtain *inverse consistent* registration (*i.e.* the registration of one set to the other provides the inverse transformation of that obtained when switching the two sets) in this *EM-ICP* framework. Such a property is very desirable, especially when building an atlas from a set of anatomical structures. First trials towards a symmetric point matching include the work by Rangarajan *et al.* who used the Sinkhorn theorem to enforce the match matrix to be doubly stochastic [3]. Most other works focused

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 594–601, 2010.

on the estimation of the transformation: Joshi & Miller [4] showed how to build a diffeomorphism between two sets of matched points, but without guaranteeing inverse consistency. In parallel, Johnson & Christensen proposed a method towards this goal, but using thin-plate splines where invertibility is not ensured [5]. These solutions are theoretically attractive but computationally redhibitory in case of large point sets, which limits their use to simple anatomical structures.

In this paper, we propose a symmetric formulation of the registration problem in an EM-ICP framework, that allows to jointly compute the forward and the backward deformation fields linking the two point sets (Section 2). Our cost function is composed of two *data attachment* terms, two *consistency* terms (to enforce both transformations to be compatible with each other) and two *regularisation* terms over both fields. We provide an iterative two-step algorithm to minimise this new criterion, in which the first step is quite similar to that of the original EM-ICP algorithm and the second step consists of two interleaved approximation problems. Using the Reproducing Kernel Hilbert Space (RKHS) theory and the Fourier analysis, we devise efficient regularisers leading to closed-form solutions based on sparse linear algebra for these two problems (Section 3). This results in an efficient algorithm allowing non-linear registration on large 3D point sets (Section 4). Finally, we give some perspectives (Section 5).

2 A Framework for a Consistent and Symmetric EM-ICP

2.1 Cost Function

Let $X = \{x_1, ..., x_N\}$ and $Y = \{y_1, ..., y_M\}$ be two point sets representing two anatomical structures to be registered. Let T^X and T^Y be respectively the backward and forward unknown transformations superposing X and Y. Let A^X (resp. A^Y) be the match matrix describing the correspondences between $T^X(X)$ and Y (resp. $T^Y(Y)$ and X). Let α and β be real positive values weighing the influence of the different terms. Following the EM-ICP framework in its energetic formulation [3], we consider the matrices A^X and A^Y as hidden variables of the problem and we design our cost function as :

$$\begin{aligned}\mathcal{E}(T^X, T^Y, A^X, A^Y) = \; & [\mathcal{E}_d(Y, T^X(X), A^X) + \mathcal{E}_d(X, T^Y(Y), A^Y) \\ & + \alpha\mathcal{E}_c(T^X \circ T^Y, I) + \alpha\mathcal{E}_c(T^Y \circ T^X, I) \\ & + \beta\mathcal{E}_r(T^X) + \beta\mathcal{E}_r(T^Y)], \end{aligned} \quad (1)$$

where:

- $\mathcal{E}_d(Y, T^X(X), A^X)$ is a data attachment term defined as:

$$\mathcal{E}_d(Y, T^X(X), A^X) = 1/N\big[\sum_{i,j} A^X_{i,j}||y_i - T^X(x_j)||^2 + \sigma^2 \sum_{i,j} A^X_{i,j} \log(A^X_{i,j})\big], \quad (2)$$

where $\sum_{i,j}$ is the sum over the points $y_i \in Y$ and $x_j \in X$, and $\sum_i A^X_{i,j} = 1 \; \forall j$. The parameter σ can be seen as the Gaussian noise variance on X and Y. In practice σ controls the fuzziness of A.

- $\mathcal{E}_c(T^Y \circ T^X, I)$ is a consistency term that measures the discrepancy between transformations T^X and T^Y. Without this term, estimations of T^X and

T^Y would be completely independent. This new term couples them and forces them to be compatible with each other. We design it as: $\mathcal{E}_c(T^Y \circ T^X, I) = 1/N \sum_{x_j \in X} ||T^Y \circ T^X(x_j) - x_j||^2$.

- $\mathcal{E}_r(T^X) = R(T^X)$ is a regularisation term penalising discontinuities of T^X.
- $\mathcal{E}_d(X, T^Y(Y), A^Y)$, $\mathcal{E}_c(T^X \circ T^Y, I)$ and $\mathcal{E}_r(T^Y)$ are built the same way.

2.2 Minimisation

The cost function (1) can be optimised by an iterative two-step algorithm that consists in minimisation over $A^X; A^Y$ and $T^X; T^Y$ in turn:

init $\tilde{T}^X$ and $\tilde{T}^Y$ as the identity function
i) $\tilde{A}^X$; $\tilde{A}^Y = \arg\min_{A^X, A^Y} \mathcal{E}(\tilde{T}^X, \tilde{T}^Y, A^X, A^Y)$
ii) $\tilde{T}^X$; $\tilde{T}^Y = \arg\min_{T^X, T^Y} \mathcal{E}(T^X, T^Y, \tilde{A}^X, \tilde{A}^Y)$

Step i) has a closed-form solution. To achieve robustness, we replace the square cost function in $\mathcal{E}_d$ by a truncated quadratic cost function. Then Step i) can be efficiently solved using a kd-tree [2]. For Step ii), we optimise iteratively the criterion with respect to each one of the two unknowns T^X and T^Y:

init estimate the optimal $\tilde{T}^X$ and $\tilde{T}^Y$ dropping the consistency term $\mathcal{E}_c$
ii.a) $T^X = \tilde{T}^X$ and $T^Y = \tilde{T}^Y$
ii.b) $\tilde{T}^X = \arg\min_{T^X} \mathcal{E}(T^X, T^Y, \tilde{A}^X, \tilde{A}^Y)$
ii.c) $\tilde{T}^Y = \arg\min_{T^Y} \mathcal{E}(T^X, T^Y, \tilde{A}^X, \tilde{A}^Y)$

Intuitively, this algorithm consists in alternatively estimating T^X as a compromise between data attachment ($\mathcal{E}_d(Y, T^X(X), A^X)$), regularisation ($\mathcal{E}_r(T^X)$) and consistency with T^Y ($\mathcal{E}_c(T^Y \circ T^X, I)$) and T^Y as a compromise between the three other symmetric terms. Implementing this last scheme requires further specification of T and R. We define the transformation T^X as the initial position plus a displacement field: $T^X(x_j) = x_j + t^X(x_j)$ and R is a regulariser on t^X (similarly for T^Y and t^Y). Then Step ii.b) can be written as:

ii.b) $\tilde{t}^X = \arg\min_{t^X} 1/N \sum_{i,j} A^X_{i,j} ||y_i - x_j - t^X(x_j)||^2 + \beta R(t^X) +$
$\alpha/N \sum_j ||\tilde{t}^Y(x_j + t^X(x_j)) + t^X(x_j)||^2 + \alpha/M \sum_i ||t^X(y_i + \tilde{t}^Y(y_i)) + \tilde{t}^Y(y_i)||^2$

Step ii.c) has a similar expression.

Due to the terms $\sum_j ||\tilde{t}^Y(x_j + t^X(x_j)) + t^X(x_j)||^2$ in step ii.a) and $\sum_i ||\tilde{t}^X(y_i + t^Y(y_i)) + t^Y(y_i)||^2$ in Step ii.b), which are somewhat redundant with their symmetric counterparts, the two problems are very intricate. Thus, similarly to what is done by Chui *et al.* [7] in a related context, we drop them, which allows to reformulate Steps ii.a) and ii.b) as two independent approximation problems:

ii.b) $\tilde{t}^X = \arg\min_{t^X} 1/N \sum_{i,j} A^X_{i,j} ||y_i - x_j - t^X(x_j)||^2$
$+\beta R(t^X) + \alpha/M \sum_i ||t^X(y_i + \tilde{t}^Y(y_i)) + \tilde{t}^Y(y_i)||^2$
ii.c) $\tilde{t}^Y = \arg\min_{t^Y} 1/M \sum_{i,j} A^Y_{i,j} ||x_i - y_j - t^Y(y_j)||^2$
$+\beta R(t^Y) + \alpha/N \sum_j ||t^Y(x_j + \tilde{t}^X(x_j)) + \tilde{t}^X(x_j)||^2$

In practice, only a few iterations are necessary to decrease importantly the criterion. Then, R could be chosen as a TPS regulariser and the two approximations problems would consist in solving linear systems of size proportional to

$N \times M$, which would be impracticable in terms of time and memory complexity for large point sets. Below, we propose an alternative efficient strategy.

3 Efficient Solutions for the Approximation Problems

3.1 Formalism

One can show that the two above mentioned problems can be restated as:

$$\tilde{f} = \arg\min_f \sum_{i,j} p_j A_{i,j} ||u_i - (v_j + f(v_j))||^2 + \beta R(f). \quad (3)$$

where $U = \{u_1, ..., u_N\}$, $V = \{v_1, ..., v_M\}$ and $\sum_i A_{i,j} = 1 \ \forall j$. The positive values (p_j) are introduced for the sake of generality and allow to consider some points of V as outliers by simply fixing $p_k = 0$ when v_k is an outlier.

First, one can show that this problem is equivalent to (considering that the derivatives vanish at the optimum and noting that $\sum_i A_{i,j} = 1, \forall j$):

$$\tilde{f} = \arg\min_f \sum_j p_j ||c_j - (v_j + f(v_j))||^2 + \beta R(f), \text{with } c_j = \sum_i A_{i,j} u_i. \quad (4)$$

This reduces the size of the problem from $\text{card}(U) \times \text{card}(V)$ to $\text{card}(V)$.

Now, we focus on building a tractable (in terms of minimisation) and powerful (in terms of reliability of the model) regulariser R. For that, we consider our problem in a space of admissible solutions $\mathcal{H}$ that we span using a positive definite kernel (pdk) k [1] : $\mathcal{H} = \{f | f(.) = \sum_{i=0}^{\infty} k(q_i, .)w_i, w_i \in I\!R^3, q_i \in I\!R^3; ||f||_{\mathcal{H}} < \infty\}^C$ where Ω^C denotes the completion of the set Ω. This space is endowed with the inner product: $< f, h >_{\mathcal{H}} = \sum_{i,j=0}^{\infty} w_i^T k(q_i, q_j) w_j$. The space $\mathcal{H}$ is a Hilbert space with reproducing kernel k (or more compactly a RKHS) [8]. Then we assume that $f \in \mathcal{H}$ and define our regulariser $R(f)$ as $||f||_{\mathcal{H}}$:

$$\tilde{f} = \arg\min_{f \in \mathcal{H}} \sum_j p_j ||c_j - (v_j + f(v_j))||^2 + \beta ||f||_{\mathcal{H}}. \quad (5)$$

One of the key advantage of RKHS is that one can show [9] that the values taken by the solution $\tilde{f}$ at the points $v_1, \ldots, v_j, \ldots, v_M$ can be expressed as $\tilde{f}(v_j) = \sum_{i=1}^{M} k(v_i, v_j) w_i$ and then formulate the last minimisation problem as:

$$(\tilde{w}) = \arg\min_{(w)} \sum_{j=1..M} p_j ||c_j - \big(v_j + \sum_{i=1..M} k(v_j, v_i) w_i\big)||^2 + \beta \sum_{i,j=1..M} w_i^T k(v_j, v_i) w_j$$

Vanishing the derivatives gives a linear system whose solution can be expressed in a closed-form as: $W = (d(P)K + \beta I)^{-1} d(P)[C - V]$, where $V = [v_1, ..., v_M]^T$, $C = [c_1, ..., c_M]^T$, $W = [w_1, ..., w_M]^T$, $K = (k(v_i, v_j)_{i,j})$ is the M by M matrix associated to kernel k and $d(P)$ is the diagonal matrix formed by the p_j values. The challenge is now to choose a kernel corresponding to a relevant regulariser.

[1] More generally, we could use a vectorial positive definite kernel $\mathbf{k}$ (in our case, $\mathbf{k}(.,.)$ would be a 3×3 matrix). By simply considering a scalar pdk, we indirectly restrict our study to vectorial pdk $\mathbf{k}$ of the form $\mathbf{k}(.,.) = k(.,.)I$. However, note that all the results of this section can be extended to vectorial pdks.

3.2 Choosing A Kernel

In order to design a suitable k, one can use an interesting relationship with Fourier-based stabilisers. Let $\forall f$ integrable, $R(f) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} \frac{|f^*(\omega)|^2}{\phi^*(||\omega||/b)} d\omega$, where * is the Fourier transform operator, $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is an integrable function and b is a real positive rescaling factor. Let $\mathcal{F} = \{f : \mathbb{R}^3 \rightarrow \mathbb{R}^3 | R(f) < \infty\}$. Interestingly, one can state that if the function $(q_i, q_j) \rightarrow \phi(||q_i - q_j||)$ is a pdk thus $\mathcal{F}$ is a RKHS whose reproducing kernel is given by $k(q_i, q_j) = b \times \phi(b \times ||q_i - q_j||)$ and such that $\forall f \in \mathcal{F}, ||f||_{\mathcal{F}} = R(f)$ [8,10]. This dual view is convenient as it allows to design a wide variety of efficient regularisers directly into the Fourier domain.

In order to design an efficient regulariser, we have to choose $\frac{1}{\phi^*}$ as a high-pass filter. This way, high frequencies of the deformation will be drastically penalised whereas low frequencies will only be penalised a little. Generally, $\phi^*_{[0,\infty]}$ is a monotonically decreasing function and the most important element that characterises its influence on the regularisation is the way it decreases that indicates the amount of penalisation with respect to frequencies. Particularly, the frequencies for which $\phi^*(||\omega||/b)$ is null are forbidden. The two parameters β and b allow to handle the regularisation properties: β is a quantitative parameter (it indicates the amount of smoothness) whereas b is more qualitative (in a way, it defines what the term "smoothness" means). Finally, note that an undesirable effect of this approach is to penalise the null frequency *i.e.* the average of the field (as $1/\phi^*(0)$ is not null). One removes this penalisation by simply ensuring that the deformation field has the same norm before and after regularisation. Figure 1 shows the influence of b and β when approximating a noisy field when choosing ϕ as the Wu kernel [8].

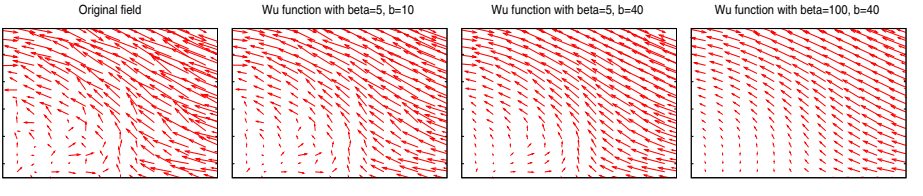


Fig. 1. Effect of parameters β and b on the approximation of a noisy field

3.3 Efficient Choices

Although we propose a closed-form solution for the approximation problem, it consists in solving a $M \times M$ system. This is can be problematic when M increases (in term of memory usage and of computational time). Suppose that we choose a compactly supported pdk (*i.e.* for $\forall x$, $\forall y$ such that $||y|| > r$; $k(x, y) = 0$), then i) $d(P)K + \beta I$ is a sparse matrix that can be computed using a kd-tree and ii) computing W consists in solving a sparse system. Some interesting compactly supported pdk corresponding to low-pass filters have been proposed in the literature (such as Wendland, Wu or Buhmann functions). Moreover, techniques to generate a wide variety of them have been proposed [8]. Alternatively one can

use a highly decreasing function and approximate it by zeroying all its values over a given threshold. We experimentally find the compact support kernel of Wu ($\phi_{2,3}$) as the one providing the best results and we will use it in the following.

4 Evaluation and Application

We use three databases of 20 surfaces each, composed of pairs of lateral ventricles of control subjects (15,000 points, segmented from T1-weighted MRI using `itksnap.org`), caudate nuclei of patients with dysphasia (2,000 points, segmented from T1-weighted MRI with an in-house software) and osseous labyrinths of modern *H. sapiens* (30,000 points, segmented from CT images with `amira.com`). We perform the following validation experiments.

Experiment 1. We choose one surface in each dataset and deform these 3 surfaces 100 times using randomly generated non-linear transformations (using TPS). Then we register the original and deformed surfaces and compute the overall residual distance between the known corresponding points and the Hausdorff distance between the surfaces. These two error measures are then averaged over the 100 simulations for each of the three surfaces.

Experiment 2. We manually select 6 landmarks on each of the 40 pairs of ventricles and labyrinths (the 20 nuclei are not used in this experiment). Then we choose one of the surfaces in each of the two datasets to be the template, and register all the 19 other surfaces to this template. We evaluate the mean residual errors on the landmarks, the Hausdorff distance between the surfaces, and we average these errors over the 19 subjects for each of the two datasets.

For both sets of experiments, we evaluate 3 strategies: asymmetric EM-ICP formulation using an order-one Tikhonov regularisation [6] (Method 1), asymmetric

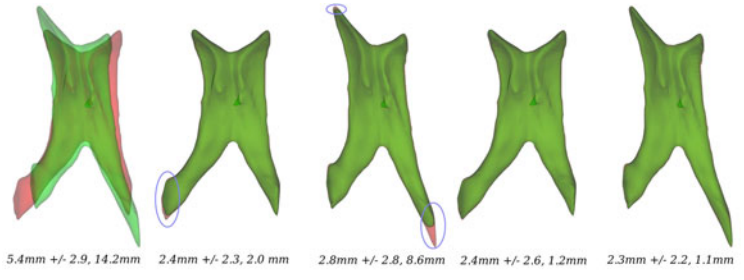


Fig. 2. Registration of lateral ventricles. From left to right: (i) two misaligned pairs of lateral ventricles A and B (from Exp. 1); (ii) A to B and (iii) B to A with the asymmetric EM-ICP with RKHS; (iv) A to B and (v) B to A with the symmetric-consistent EM-ICP with RKHS. Point-to-point and Hausdorff registration errors are given below each figure. The asymmetric formulation leads to registration errors close to the horns. The location of these errors depends on what surface is used as the template. With the symmetric consistent formulation, no order-dependent registration error is visible, and the overall registration quality is visually and quantitatively improved.

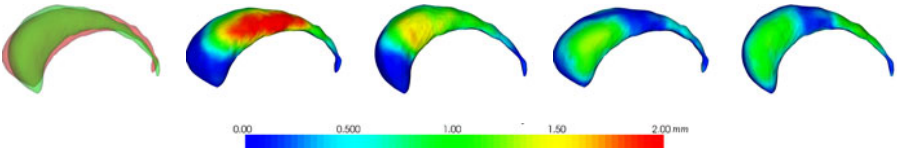


Fig. 3. Registration of caudate nuclei. Same display as in Fig. 2, but with caudate nuclei and a color mapping of the point-to-point errors. We draw the same conclusions as in Fig. 2.

Table 1. Average registration errors for the 3 tested methods and the two experiments on the 3 datasets. The first error is the mean point-to-point (for Exp. 1) or landmark-to-landmark (Exp. 2) error, and the second error is the Hausdorff distance. Both errors are averaged over the 100 (Exp. 1) or 19 (Exp. 2) registrations. Method 3 (symmetric-consistent EM-ICP with RKHS regularisation) gives the lowest error for each experiment and dataset.

	Experiment 1			**Experiment 2**		
	Method 1	Method 2	Method 3	Method 1	Method 2	Method 3
nuclei	2.26/1.87	0.73/1.76	0.64/1.37	-	-	-
ventricles	2.15/7.80	1.55/5.15	1.53/1.48	2.21/3.41	1.84/2.23	1.67/1.01
labyrinths	0.29/0.85	0.23/0.75	0.21/0.23	0.32/0.65	0.28/0.62	0.24/0.19

formulation using RKHS-based regularisation (Method 2) and symmetric-consistent formulation using RKHS (Method 3). The results are displayed on Tab. 1. Fig. 2 and Fig. 3 illustrate the added value of the symmetric consistent RKHS formulation compared to the asymmetric one on ventricles and nuclei.

The parameters for Method 3 are initialised as: $\alpha = 8$, $\beta = 400$, $\sigma^2 = 20 \times S$, $b = 250 \times S$ where S is the size (in metre) of the object; β, σ^2, b are then decreased throughout the algorithm until they reach the respective values of 20, $5 \times S$ and $150 \times S$. The same is done for Method 1 for β and σ^2 and for Method 2 for β, σ^2 and b. The run time to register point sets of 5,000 points is 2 min for Method 1, 5 min for Method 2 and 12 min for Method 3 on a standard PC.

Application to statistical shape modelling. Two major tracks have been followed in the literature to extend the EM-ICP formalism to build statistical shape models [7,12]. By using our contributions, these methods can be made symmetric and consistent (allowing to reduce the bias during the computation

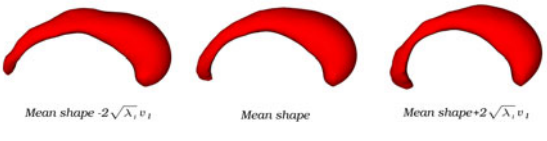


Fig. 4. Mean shape and first mode of variation ($\pm$ 2 $\sqrt{\lambda_1}$) on 20 caudate nuclei. The first mode can be interpreted as a bending.

of the mean shape) and one can drastically increase the size of the input data. Fig. 4 shows the mean shape and the first mode of variation for 20 caudate nuclei by adapting the method of Hufnagel *et al.* [12].

5 Conclusion

Our contributions are twofold. First, we proposed a symmetric and consistent EM-ICP framework where we enforce coherency between backward and forward deformations. Note that none of these transformations is guaranteed to be invertible, though. Second, we designed new efficient regularisers for EM-ICP based registration. These new regularisers are expressed within the RKHS formalism which leads to computationally attractive solutions (especially when using kernels with compact support). A particular class of kernels provides regularisers with a simple interpretation in the frequency domain. Overall, our algorithm allows to perform registration on large data sets. Future works will include i) comparing our regulariser with others such as Coherent Point Drift [11] or TPS; ii) comparing our statistical shape modelling with other techniques [7,12].

Acknowledgements. We thank José Braga for the labyrinths data and Aline Carsin for segmenting the caudate nuclei.

References

1. Besl, P., McKay, N.: A method for registration of 3-D shapes. IEEE PAMI 14(2), 239–256 (1992)
2. 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)
3. Chui, H., Rangarajan, A.: A Feature Registration Framework Using Mixture Models. IEEE MMBIA, 190–198 (2000)
4. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. IEEE TMI 9(8), 1357–1370 (2000)
5. Johnson, H., Christensen, G.: Consistent landmark and intensity-based image registration. IEEE TMI 21(5), 450–461 (2002)
6. Combès, B., Prima, S.: Setting priors and enforcing constraints on matches for nonlinear registration of meshes. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 175–183. Springer, Heidelberg (2009)
7. Chui, H., et al.: Unsupervised learning of an atlas from unlabeled point-sets. IEEE PAMI 26, 160–172 (2004)
8. Wendland, H.: Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics (2005)
9. Schölkopf, B., Herbrich, R., Smola, A.J.: A generalized representer theorem. In: Annual Conference on Computational Learning Theory, pp. 416–426 (2001)
10. Sidlofov, T.: Existence and uniqueness of minimization problems with Fourier based stabilizers. Compstat, Prague (2004)
11. Myronenko, A., et al.: Non-rigid point set registration: Coherent Point Drift. In: Advances in Neural Information Processing Systems, pp. 1009–1016 (2007)
12. Hufnagel, H., et al.: Generation of a Statistical Shape Model with Probabilistic Point Correspondences and EM-ICP. IJCARS 2(5), 265–273 (2008)

Image Registration Driven by Combined Probabilistic and Geometric Descriptors*

Linh Ha[1], Marcel Prastawa[1], Guido Gerig[1], John H. Gilmore[2], Cláudio T. Silva[1], and Sarang Joshi[1]

[1] Scientific Computing and Imaging Institute, University of Utah
[2] Department of Psychiatry, University of North Carolina

Abstract. Deformable image registration in the presence of considerable contrast differences and large-scale size and shape changes represents a significant challenge for image registration. A representative driving application is the study of early brain development in neuroimaging, which requires co-registration of images of the same subject across time or building 4-D population atlases. Growth during the first few years of development involves significant changes in size and shape of anatomical structures but also rapid changes in tissue properties due to myelination and structuring that are reflected in the multi-modal Magnetic Resonance (MR) contrast measurements. We propose a new registration method that generates a mapping between brain anatomies represented as a multi-compartment model of tissue class posterior images and geometries. We transform intensity patterns into combined probabilistic and geometric descriptors that drive the matching in a diffeomorphic framework, where distances between geometries are represented using currents which does not require geometric correspondence. We show preliminary results on the registrations of neonatal brain MRIs to two-year old infant MRIs using class posteriors and surface boundaries of structures undergoing major changes. Quantitative validation demonstrates that our proposed method generates registrations that better preserve the consistency of anatomical structures over time.

1 Introduction

Image registration is a basic task in defining a standard space for analyzing populations that change over time, which is essential to determine development in normal growth and neurological disorders. The growth process can involve large-scale size and shape changes, as well as changes in tissue properties and appearance. These factors pose significant challenges in image registration, as image intensities need to be interpreted differently at different stages. A strong example is the human brain at early development stages.

A driving research question is to determine the process of white matter myelination, which manifests as two distinct white matter appearance patterns primarily during the first year of development. Other clinical research questions are related

* Supported by NIH grants 5R01-EB007688, P41-RR023953, Conte Center MH064065, and NSF grant CNS-0751152.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 602–609, 2010.

to finding a link between cognitive development and the rapid, locally varying growth of specific anatomical structures. To approach these questions, a robust registration method is necessary for mapping longitudinal brain MRI to a reference space so that we can perform reliable analysis of the tissue property changes reflected in the MR measurements. Knickmeyer *et al.*[1] showed that the total brain volume grows by 100% the first year and 15% the second year, whereas the cerebellum shows 220% volume growth for the first and another 15% for the second year, indicating the very different growth rates of different anatomical structures. Through regression on shape representations, Datar *et al.*[2] illustrated that the rapid volume changes are also paralleled by significant shape changes which describe the dynamic pattern of localized, nonlinear growth. These challenges require a method that does not rely on raw intensity measurements, while also being capable of estimating large structural deformations. Xue *et al.*[3] addressed these issues by proposing a registration scheme for neonatal brains by registering inflated cortical surfaces extracted from the MRI.

We propose a new registration framework for longitudinal brain MRI that makes use of the underlying anatomies, which are represented by both class posteriors and boundary surfaces. This framework is able to match internal anatomical regions and simultaneously preserving a consistent mapping for the boundaries of relevant anatomical objects. We show results of registering neonatal brain MRI to 2-year old brain MRI of the same subjects obtained in a longitudinal neuroimaging study. Our method consistently provides transformations that better preserve time-varying structures than obtained by intensity-only registration.

2 Registration between Anatomies

We propose a registration method that makes use of the underlying anatomy in the MR images. Fig. 1 shows an overview of the registration process. We begin by extracting anatomical descriptors from the images, followed by computing a transformation that minimizes the distance between the anatomical descriptors.

2.1 Anatomical Descriptors

We represent brain anatomy as a multi-compartment model of tissue classes and manifolds. We associate each position x with a vector of tissue probability densities. In a given anatomy, we capture the underlying structures by estimating, for each image, the class posterior mass functions associated with each of the classes. Given Ω as the underlying coordinate system of the brain anatomies, each anatomy $\mathcal{A}_{i=1,\cdots,N}$ is represented as

$$\mathcal{A}_i = \{p_{i,c=1}(x), \cdots, p_{i,c=N_c}(x), \mathcal{M}_{i,j=1}(2), \cdots, \mathcal{M}_{i,j=N_s}(2) \subset \Omega\} \qquad (1)$$

where N_c is the number of probability images, N_s is the number of surfaces, $p_c(x)$ is the class posterior for tissue c at location x, and $\mathcal{M}_j(2)$ are 2-dimensional submanifolds of Ω (surfaces).

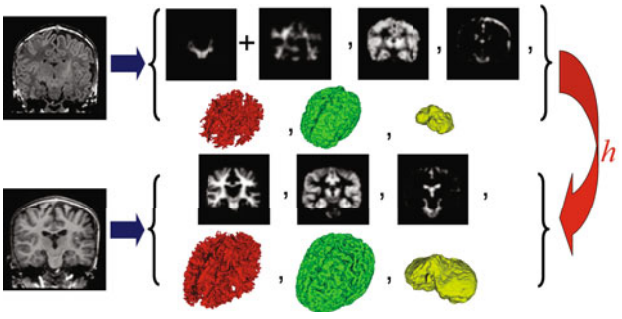


Fig. 1. Overview of the proposed registration method that can handle large deformations and different contrast properties, applied to mapping brain MRI of neonates to 2-year olds. We segment the brain MRIs and then extract equivalent anatomical descriptors by merging the two different white matter types present in neonates. The probabilistic and geometric anatomical descriptors are then used to compute the transformation h that minimizes the distance between the class posterior images, as well as the distance between surfaces represented as currents.

The classification of brain MR images with mature white matter structures into class posteriors are well studied. We extract the posteriors from 2-year old brain MR images using the segmentation method proposed by van Leemput *et al.*[4]. The method generates posterior probabilities for white matter (wm), gray matter (gm), and cerebrospinal fluid csf. These probabilities can then be used to generate surfaces from the maximum a posteriori tissue label maps.

The classification of neonatal brain MR images is challenging as the white matter structure undergoes myelination, where the fibers are being covered in myelin sheathes. Several have proposed methods that make use of prior information from an atlas or template that takes into account the special white matter appearance due to myelination [5] [6]. We use the method described by Prastawa *et al.*[6] for extracting the tissue class posteriors of neonatal brain MRI which includes for myelinated wm, non-myelinated wm, gm, and csf. These can then be used to create an equivalent anatomy to the 2-year old brain by combining the two white matter class probabilities and surfaces.

For the results in this paper, we use the probabilities $\{p_{wm}(x), p_{gm}(x), p_{csf}(x)\}$ and we use the surfaces of white matter, gray matter, and cerebellum. The cerebellum surfaces are generated from semi-automated segmentations that are obtained by affinely registering a template image followed by a supervised level set segmentation. The cerebellum has a significant role in motor function and it is explicitly modeled as it undergoes the most rapid volume change during the first year of development and thus presents a localized large-scale deformation.

2.2 Registration Formulation

Given two anatomies $\mathcal{A}_1$ and $\mathcal{A}_2$, the registration problem can be formulated as an estimation problem for the transformation h that minimizes

$$\hat{h} = \arg\min_{h} E(h \cdot \mathcal{A}_1, \mathcal{A}_2)^2 + D(h, e)^2 \tag{2}$$

where $h \cdot \mathcal{A}_1$ is the transformed anatomy, $E(\cdot, \cdot)$ is a metric between anatomies and $D(\cdot, e)$ is a metric on a group of transformations that penalizes deviations from the identity transformation e.

We define distance between anatomies E by defining a norm on an anatomy as a combination of the L^2 norm on the class posteriors and a Reproducing Kernel Hilbert space norm on the manifolds defined as "currents" through Glaunes [7]. The currents norm does not require geometric correspondence and thus can be used to register manifolds with different resolutions. For an oriented surface $\mathcal{M}(2)$ in R^3 the norm $[\mathcal{M}(2)]$ is the vector valued Borel measure corresponding to the collection of unit-normal vectors to $\mathcal{M}(2)$, distributed with density equal to the element of surface area ds and can be written as $\eta(x)ds(x)$, where $\eta(x)$ is the unit normal and $ds(x)$ is the surface measure at point x. When $\mathcal{M}(2)$ is a discrete triangular mesh with N_f faces, a good approximation of the norm can be computed by replacing $[\mathcal{M}(2)]$ by a sum of vector-valued Dirac masses:

$$\|[\mathcal{M}(2)]\|_k^2 = \sum_{f=1}^{N_f} \sum_{f'=1}^{N_f} \langle \eta(f), \eta(f') \rangle \, k(c(f), c(f')), \tag{3}$$

where $k(\cdot, \cdot)$ is a shift-invariant kernel (e.g., Gaussian or Cauchy), N_f is the number of faces of the triangulation, and for any face f, $c(f)$ is its center and $\eta(f)$ its normal vector with the length capturing the area of each triangle.

Having defined the norm on probability images and surfaces, the dissimilarity metric between anatomies $\left\| [\mathcal{A}_1] - [\mathcal{A}_2] \right\|_k^2$ is given by:

$$\sum_{c=1}^{N_c} \int_{\Omega} |p_{1,c}(x) - p_{2,c}(x)|^2 dx + \sum_{j=1}^{N_s} \|[\mathcal{M}_{1,j}(2) \cup (-\mathcal{M}_{2,j}(2))]\|_k^2 \tag{4}$$

where the distance between two surface currents $\|[\mathcal{M}_{1,j}(2) - \mathcal{M}_{2,j}(2)]\|_k = \|[\mathcal{M}_1(2) \cup (-\mathcal{M}_2(2))]\|_k$ is computed as the norm of the union between surface $\mathcal{M}_1(2)$ and surface $\mathcal{M}_2(2)$ with negative measures.

We use the large deformation framework [8] that generates dense deformation maps in R^d by integrating time-dependent velocity fields. The flow equation is given by $\frac{\partial h^v(t,x)}{\partial t} = v(t, h^v(t, x))$, with $h(0, x) = x$, and we define $h(x) := h^v(1, x)$, which is a one-to-one map in $\mathbb{R}^d$ (diffeomorphism). We define an energy functional that ensures the regularity of the transformations on the velocity fields: $\|v(t, \cdot)\|_V^2 = \int_{\mathbb{R}^d} \langle Lv(t, x), Lv(t, x) \rangle \, dx$, where L is a differential operator acting on vector fields. This energy also defines a distance in the group of diffeomorphisms:

$$D^2(h, e) = \inf_{v, p^v(1,\cdot)=h} \int_0^1 \|Lv(t)\|_V^2 \, dt. \tag{5}$$

The registration optimizations in this paper are performed using a greedy approach by iteratively performing gradient descent on velocity fields and updating

the transformations via an Euler integration of the O.D.E. At each iteration of the algorithm the velocity field is calculated by solving the p.d.e $Lv = F(h)$ where $F(h)$ is the variation of $\left\|[h \cdot \mathcal{A}_1] - [\mathcal{A}_2]\right\|_k^2$ with respect to h. This variation is a combination of the variation of the L^2 norm on the class posteriors and of the currents norm; computed using the gradient $\frac{\partial \|[\mathcal{M}(2)]\|_k^2}{\partial x_r}$:

$$\sum_{f|x_r \in f} \left[\frac{\partial \eta(f)}{\partial x_r}\right] \sum_{f'=1}^{N_f} k(c(f'), c(f))\eta(f') + \frac{2}{3} \sum_{f'=1}^{N_f} \frac{\partial k(c(f), c(f'))}{\partial c(f)} \eta(f')^t \eta(f) \quad (6)$$

given that points $\{x_r, x_s, x_t\}$ form the triangular face f and its center $c(f) = \frac{x_r + x_s + x_t}{3}$ and its area-weighted normal $\eta(f) = \frac{1}{2}(x_s - x_r) \otimes (x_t - x_r)$.

2.3 Efficient Norm Computation Using Particle Mesh Approximation on GPU

The major challenge of implementing the currents norm (Eq. 3) for realistic brain surfaces is the high computational cost to compute the dissimilarity metric of all pairs of surface elements, which is $O(N_f^2)$ where N_f is the number of faces (can be up to millions). For computational tractability, Durrleman *et al.* [9] used a sparse representation of the surface based on matching pursuit algorithm. An efficient framework based on the standard fast Gauss transform [10] requires the construction and maintenance of the kd-tree structure on the fly, which is slow on the current CPU model. Our method, however, exploits the Particle-Mesh approximation to reduce the complexity to $M \log M$ where M is the volume size of the embedded grid. The grid size is chosen as the size of the input images to limit the approximation error to the same order of matching term for the class posteriors. This approximation have been extensively studied in the cosmological N-body simulation literature (see Hockney and Eastwood [11] for details). Particle mesh framework shares the same computational grid with the class posteriors, moving the computation to the grid level and enabling an efficient parallel computation.

Even with the particle mesh approximation of the norm computation, the total complexity of the method is still very high. On a high-end workstation with 8-CPU cores, a highly optimized multi-threaded implementation in C++ takes several hours for one matching pair, hence can not be used for parameter exploration and real-time analysis. Fortunately, this computational cost can be amortized using the massive parallel power and bandwidth provided by the graphics processing unit (GPU). Based on the GPU framework by Ha *et al.*[12], we developed an implementation that runs entirely on the GPU. The main benefit of a GPU implementation is the ability to exploit parallel efficiency of regular grid presentation.

Computing the currents u and gradient between a surface with 160535 triangular faces and another with 127043 faces takes approximately 504 seconds on an AMD Phenom II X4 955 CPU, while it takes 0.33 seconds on an NVIDIA

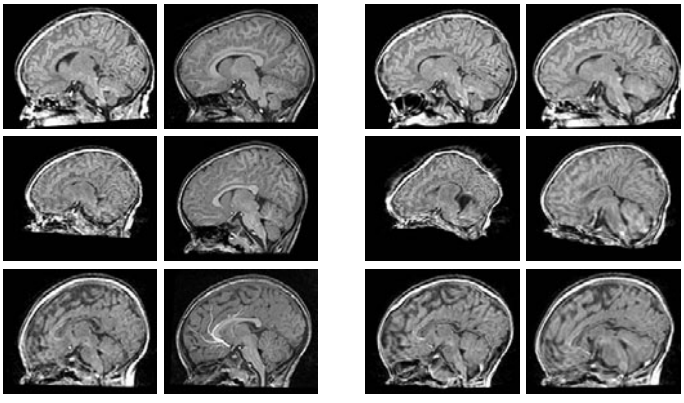


Fig. 2. Registration results of neonates mapped to two-year olds. From left to right: (a) neonatal T1 image after affine registration, (b) reference T1 image at 2-years, followed by (c) neonatal T1 after deformable mutual information registration using B-splines, and (d) after combined probabilistic and geometric registration. From top to bottom: subject 0012, 0102, 0106.

GTX 260 GPU. The speed gain is in order of three magnitudes over the equivalent CPU implementation using particle mesh, while the computing time for the exact norm on CPU is difficult to measure since it takes significantly longer. The proposed algorithm typically converges in 1000 iterations, so on average it takes less than eight minutes to register two anatomies. This allows us to perform parameter exploration and real-time analysis on a single desktop with commodity GPU hardware. The efficiency of the GPU method also provides an opportunity to apply the algorithm for high quality atlas formation using our framework on a GPU cluster, which gives us the ability to perform statistical tests that are previously impossible due to excessive time requirements.

3 Results

We have applied the registration method for mapping neonatal MRI scans to two-year MRI scans of the same subjects in ten datasets. The datasets are taken from an ongoing longitudinal neuroimaging study with scans acquired at approximately two weeks, one year, and two years of age. Due to rapid early brain development, each longitudinal MR scan shows significant changes in brain size and in tissue properties. For comparison, we also applied the standard intensity based deformable registration using mutual information (MI) metric and B-spline transformation proposed by Rueckert *et al.* [13] which has been applied for registering 1-year old and 2-year old infants. The T1 weighted images before and after registration using the different approaches for the first three subjects are shown in Fig. 2.

A quantitative study of the performance of the registration method is performed by measuring the overlap between the transformed segmentation maps

	Subject	0012	0102	0106	0121	0130	0146	0156	0174	0177	0180
White matter	MI	0.329	0.155	0.281	0.384	0.379	0.230	0.257	0.300	0.350	0.301
	P+G	0.497	0.397	0.442	0.453	0.482	0.414	0.461	0.440	0.478	0.442
Cerebellum	MI	0.755	0.212	0.588	0.515	0.732	0.820	0.713	0.569	0.631	0.777
	P+G	0.881	0.821	0.875	0.878	0.858	0.899	0.907	0.885	0.896	0.892

Fig. 3. Overlap measures comparing the registered segmentation maps against the reference segmentation maps for the white matter and cerebellum structure, obtained through deformable mutual information registration (MI) and our proposed method (P+G)

Subject	0012	0102	0106	0121	0130	0146	0156	0174	0177	0180
MI on CPU	92	63	103	92	101	112	106	99	91	96
P+G on GPU	9	8	8	8	8	7	9	8	7	7

Fig. 4. Time elapsed, in minutes, for registration using deformable mutual information on CPU (MI) and our proposed approach (P+G) on GPU with 1000 iterations of gradient descent

of neonates to the segmentation maps of two-year olds. Since we consider the segmentation maps at two years of age to be the standard, we use the following overlap metric:

$$Overlap(h \cdot S0, S2) = \frac{|h \cdot S0 \cap S2|}{|S2|} \tag{7}$$

where $h \cdot S0$ is the transformed neonate segmentation map, $S2$ is the reference two-year segmentation map, and $|\cdot|$ indicates the volume of a binary map. Please note that this metric gives considerably lower values than the standard Dice coefficient. Fig. 3 shows the quantitative analysis for the white matter and cerebellum segmentation maps. Registration using both probabilistic and geometric descriptors provides consistently better results and are generally more stable for the structures of interest. In particular, our method better preserves the shape of the cerebellum, which has weak intensity boundaries in regions where it touches the cerebrum and thus cannot be registered properly using only image based information. Another significant challenge is that the cerebellar growth is distinctly different from the growth of neighboring structures. Registrations using our approach on the GPU takes 8 minutes on average, while registration on the CPU using mutual information metric and B-spline transformation takes 100 minutes on average. Detailed time measures are listed in Fig. 4. We have also performed validation using only the probabilistic descriptor, which generates results that are less accurate compared to our method (particularly for the cerebellum) while more accurate than image registration using MI.

4 Conclusions

We have proposed a registration framework that makes use of the probabilistic and geometric structures of anatomies embedded in the images. This allows us to

enforce matching of important anatomical features represented as regional class posteriors and tissue boundaries. Our framework allows us to register images with different contrast properties by using equivalent anatomical representations, and we have demonstrated results for registering brain MRIs with different white matter appearances at early stages of growth. The overlap validation measures in Fig. 3 show that geometric constraints, particularly for the cerebellum, is crucial for registering structures undergoing significant growth changes. In the future, we plan to apply this framework in early neurodevelopmental studies for analyzing the effects of neurological disorders such as autism and Fragile X syndrome. The proposed registration framework is generic and independent of the application domain, it can thus be applied to any registration where one encounters large-scale deformation and different appearance patterns.

References

1. Knickmeyer, R.C., Gouttard, S., Kang, C., Evans, D., Wilber, K., Smith, J.K., Hamer, R.M., Lin, W., Gerig, G., Gilmore, J.H.: A structural MRI study of human brain development from birth to 2 years. J. Neurosci. 28, 12176–12182 (2008)
2. Datar, M., Cates, J., Fletcher, P., Gouttard, S., Gerig, G., Whitaker, R.: Particle based shape regression of open surfaces with applications to developmental neuroimaging. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5762, pp. 167–174. Springer, Heidelberg (2009)
3. Xue, H., Srinivasan, L., Jiang, S., Rutherford, M.A., Edwards, A.D., Rueckert, D., Hajnal, J.V.: Longitudinal cortical registration for developing neonates. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 127–135. Springer, Heidelberg (2007)
4. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE Trans. Medical Imaging 18, 897–908 (1999)
5. Warfield, S.K., Kaus, M., Jolesz, F.A., Kikinis, R.: Adaptive, template moderated, spatially varying statistical classification. MedIA 4(1), 43–55 (2000)
6. Prastawa, M., Gilmore, J.H., Lin, W., Gerig, G.: Automatic segmentation of mr images of the developing newborn brain. MedIA 9(5), 457–466 (2005)
7. Glaunes, J., Trouvé, A., Younes, L.: Diffeomorphic matching of distributions: a new approach for unlabelled point-sets and sub-manifolds matching. In: CVPR (2004)
8. Miller, M.I., Younes, L.: Group action, diffeomorphism and matching: a general framework. Int. J. Comp. Vis. 41, 61–84 (2001)
9. Durrleman, S., Pennec, X., Trouvé, A., Ayache, N.: Sparse approximations 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)
10. Greengard, L., Strain, J.: The fast gauss transform. SIAM J. Sci. Stat. Comput. 12(1), 79–94 (1991)
11. Hockney, R.W., Eastwood, J.W.: Computer Simulation Using Particles. Taylor & Francis, Abington (1989)
12. Ha, L., Kruger, J., Fletcher, P.T., Joshi, S., Silva, C.T.: Fast parallel unbiased diffeomorphic atlas construction on multi-graphics processing units. In: EGPGV (2009)
13. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations. IEEE Trans. Med. Imag. 18(8), 712–721 (1999)

Simultaneous Fine and Coarse Diffeomorphic Registration: Application to Atrophy Measurement in Alzheimer's Disease

Laurent Risser[1,2], François-Xavier Vialard[1], Robin Wolz[2], Darryl D. Holm[1], and Daniel Rueckert[2]

[1] Institute for Mathematical Science, Imperial College London, 53 Prince's Gate, SW7 2PG, London, UK*
[2] Visual Information Processing, Imperial College London, Huxley Building, Department of Computing, SW7 2BZ, London, UK

Abstract. In this paper, we present a fine and coarse approach for the multiscale registration of 3D medical images using Large Deformation Diffeomorphic Metric Mapping (LDDMM). This approach has particularly interesting properties since it estimates large, smooth and invertible optimal deformations having a rich descriptive power for the quantification of temporal changes in the images. First, we show the importance of the smoothing kernel and its influence on the final solution. We then propose a new strategy for the spatial regularization of the deformations, which uses simultaneously fine and coarse smoothing kernels. We have evaluated the approach on both 2D synthetic images as well as on 3D MR longitudinal images out of the Alzheimer's Disease Neuroimaging Initiative (ADNI) study. Results highlight the regularizing properties of our approach for the registration of complex shapes. More importantly, the results also demonstrate its ability to measure shape variations at several scales simultaneously while keeping the desirable properties of LDDMM. This opens new perspectives for clinical applications.

1 Introduction

Recent years have seen the development of new non-rigid registration techniques allowing large diffeomorphic deformations. Such deformations are by definition smooth and invertible, properties that are highly desirable in image registration. In particular, the framework of Large Deformation Diffeomorphic Metric Mapping (LDDMM) [4,10] encodes the optimal flows of deformation in time-dependent velocity fields, that are geodesic. The optimal flow between two registered images is then the shortest path between the images according to a metric regularizing the deformation. The LDDMM approach has therefore convenient properties for the statistical comparison of images and the creation of atlases.

* We thank the Imperial College Strategic Initiative Fund for partial support. DDH was also partially supported by the Royal Society of London Wolfson Research Merit Award. We are finally thankful to Martins Bruveris for constructive discussions.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 610–617, 2010.

A practical implementation of the this framework for image registration has been proposed in [4] and successfully applied to inter-subject local shape comparison and atlas creation. However, although it was designed to allow large deformations, its practical use in real medical images is often limited to relatively small deformations. Alternatives approaches have been proposed in the literature: For instance, a symmetric interpretation of [4] using cross correlation to measure the similarity between source and target images was proposed in [3]. Other approaches, allowing multimodal registration for atlas creation [9] or using Navier-Stokes equation [5] were also proposed. Importantly, a class of diffeomorphic registration techniques, using stationary velocity fields, have emerged recently. Such parameterizations are efficient in terms of memory requirements and computational time while providing registrations similar to those obtained using the LDDMM time-dependent velocity fields [1,2,8,13]. However, although these alternatives also estimate diffeomorphic deformations, none of them has been explicitly designed to estimate geodesic transformations.

In an attempt to improve the practical usability of LDDMM for 3D medical imaging, we address in this paper a fundamental aspect of the LDDMM framework: the choice of the metric. The metric is indeed directly related to the smoothing kernel of the deformations and therefore controls their spatial regularization. In practice, small kernels will favor deformations that match local details and large kernels deformations that match global structures. Since 3D medical images may contain complex anatomical structures, with features varying at different scales, multiscale registration techniques are of high interest. However, the classical coarse to fine strategies based on Gaussian kernels are theoretically (and practically) not appropriate in the context of LDDMM since the registration at the final scale only reflects the shape variations at the finest scale. The key contribution here is then the development of a theoretically well-justified *simultaneous fine and coarse registration strategy* in the LDDMM framework. The idea underlying this paper is related to the strategy proposed in [7] which consists in registering images at several scales simultaneously.

In Sec. 2, we present our multiscale extension of [4]. Importantly, we show that this extension is particularly suitable to describe the deformations of shapes whose features exist across several scales. We then discuss in Sec. 3, its behaviour for registrations requiring large deformations and its ability to detect shape variations at several scales simultaneously. We finally present the usability of our approach on medical images by comparing 3D MR longitudinal images out of the Alzheimer's Disease Neuroimaging Initiative (ADNI) study.

2 Registration Using Multi-kernel LDDMM

2.1 Registration Using LDDMM

We first give a brief summary of the classical LDDMM registration algorithm [4]. Let I_S be a source image, defined on a spatial domain Ω, and registered on a target image I_T through the time-dependent diffeomorphic transformation ϕ_t of Ω, $t \in [0, 1]$. This transformation is defined by a time dependent velocity field v_t,

$t \in [0,1]$ as follows: $\partial_t \phi_t = v_t(\phi_t)$, where $\phi_0 = Id$ and $t \in [0,1]$. For notational convenience we introduce $\phi_{t,s} \doteq \phi_s \circ \phi_t^{-1}$. The registration problem then consists in finding the velocity fields v_t that minimizes the sum of a similarity and a deformation energy:

$$E(v) = \int_0^1 \frac{1}{2}||v_t||_V^2 \mathrm{d}t + \frac{1}{2}||I_S \circ \phi_1^{-1} - I_T||_{L^2}^2 \,. \tag{1}$$

The second term measures the similarity between the deformed source and target images, here, the sum of squared differences of the intensities in both images. To define the energy of the deformation, the time dependent velocity field v is assumed to lie in $L^2([0,1],V)$, where the Hilbert space V is expressed by a smooth matrix valued kernel $k(.,.)$ describing the velocity fields that can be used for the registration. Even though there is a wide family of available kernels associated with V, most approaches use Gaussian kernels:

$$K(x) = (2\pi)^{-d/2}|\Sigma|^{-1/2} \exp\left(-\frac{1}{2}x^T \Sigma^{-1} x\right), \tag{2}$$

where Σ is the covariance matrix of the Gaussian kernel. In our work, we assume isotropic covariances, *i.e.* $\Sigma = \sigma Id_{\mathbb{R}^d}$, where the key parameter σ controls the level of the spatial correlation of the deformations. The minimization algorithm is described hereafter. We denote $J_t^S = I_S \circ \phi_{t,0}$, $J_t^T = I_T \circ \phi_{1,t}$ and $|D\phi_{t,1}|$ the Jacobian of $\phi_{t,1}$ at time t. The minimization of the variational problem of eq. 1 is performed by using a steepest gradient descent approach. This involves the iterative use of the gradient of E in $L^2([0,1],V)$, $\forall t$:

$$\nabla_v E_t = v_t - K \star \left(|D\phi_{t,1}^v| \nabla J_t^S (J_t^S - J_t^T)\right), \tag{3}$$

where $\star$ denotes the convolution operator. The velocity field is then updated by computing $v^{k+1} = v^k - \epsilon \nabla_{v_{t_j}^k} E$, where ϵ controls the step size during the gradient descent. After convergence towards the minimum energy solution, the resulting time dependent diffeomorphism is a geodesic path in the group of diffeomorphisms for which the associated velocity field satisfies the Euler-Lagrange equation.

2.2 Multi-kernel LDDMM

As discussed in introduction, the comparison of shapes can only be performed at a single scale when using [4] with classical Gaussian kernels. To perform the registration at several scales simultaneously, we simply define the kernel K as the weighted sum of N Gaussian kernels K_n of different sizes as follows:

$$K(x) = \sum_{n=1}^{N} K_n(x) = \sum_{n=1}^{N} a_n (2\pi)^{-3/2} |\Sigma_n|^{-1/2} \exp\left(-\frac{1}{2}x^T \Sigma_n^{-1} x\right), \tag{4}$$

where Σ_n and a_n are respectively the covariance matrix and the weight of the n^{th} Gaussian function. As in eq. 2, each Σ_n is only defined by a characteristic

length σ_n: $\Sigma_n = \sigma_n Id_{\mathbb{R}^d}$. Eq. 4 allows to construct a wide range of kernels with several scales of interest while preserving the promising statistical properties of the LDDMM. The optimization is then performed simultaneously at several scales so that complex shapes can be compared in this framework. Note that the choice of the weights $(a_n)_{n\in[1,N]}$ (discussed in more detail in sec. 2.4) is a key issue here, since it controls the influence of the deformations at different scales.

2.3 Separating the Contribution of Each Kernel

Interestingly, our multi-kernel extension of [4] can be re-written to distinguish the contribution of each kernel K_n in the total deformation. A velocity field v_n is associated with each kernel K_n. The energy gradient described in eq. 3 becomes:

$$\nabla_{v_n} E_t = v_{n,t} - K_n \star \left(|D\phi^v_{t,1}| \nabla J^S_t (J^S_t - J^T_t) \right), \quad \forall n, \tag{5}$$

and the velocity field is updated by using:

$$v^{k+1}_n = v^k_n - \epsilon \nabla_{v^k_{n,t_j}} E, \quad \forall n. \tag{6}$$

This formulation of the gradient descent is equivalent to the original one and provides the same deformations. Using the equations 5 and 6, the sum of the fields v_n is indeed equal to the total velocity field: $v = \sum_{n=1}^{N} v_n$. The deformation of the source image can also be computed using $\partial_t \phi_t = \sum_{n=1}^{N} v_{n,t}(\phi_t)$ and $\phi_0 = Id$. Observing the contribution of a kernel K_n at a point x of the deformed image is then naturally done by observing the amplitude of the deformation (AOD) generated by $v_{n,t}$ along $\phi_t(x)$, $t \in [0,1]$: $AOD_n(x) = \int_0^1 |v_{n,t}(\phi_t \circ x)| \mathrm{d}t$. Obviously, using v_t instead of $v_{n,t}$ in this equation gives the length of the total deformation from x. As seen in section 3.2, the AOD is an interesting descriptor of the local deformations for shapes whose features exist across several scales.

2.4 Weight of the Kernels

Our multiscale strategy depends on a set of parameters a_n, $n \in [1, N]$, each of them controlling the weight of the deformations at scale n. These weights reflect indirectly the amplitude of the deformations expected between the different scales. We introduce the apparent weights a'_n, $n \in [1, N]$ that give to the user an intuitive control of the deformation at each scale. Importantly, similar values of a'_n induce deformations having similar amplitudes at each scale n. The weights are then computed using $a_n = a'_n / g_n(K_n, I_S, I_T)$ where g_n is the typical amplitude of the updates in eq. 6 between I_S and I_T, if K_n has a weight equal to 1. For each K_n, the value of g_n can be estimated by observing the maximum update of the velocity field v in a pre-iteration of the registration of I_S on I_T using only the kernel K_n with $a_n = 1$. This simple technique is shown to work efficiently both on 2D synthetic images and 3D medical images. More importantly, for images of the same type (*e.g.* MR brain images with a fixed acquisition protocol) and the same kernels K_n, the values of g_n were observed stable. The method can then be used for an atlas creation or multiple comparisons by systematically using the same kernels with same weights.

3 Evaluation

3.1 Application to 2D Synthetic Images

As discussed above, the kernel controls the spatial behavior of the deformations. For an isotropic Gaussian kernel with standard deviation σ, the parameter σ is the characteristic length defining the scale at which the registration is performed. To illustrate the behavior of the deformation as a function of the kernel, we show in Fig. 1 an example of deformation in which a square (I_S) is registered to a rectangle containing an indentation (I_T) using a small kernel ($\sigma = 1.5$ pixels), a large kernel ($\sigma = 5$ pixels) and the sum of kernels with the same apparent weights. Note that the deformation grid obtained using the small kernel seems to highlight a non-invertible deformation. Refining this grid shows that the deformation is actually invertible but with very high Jacobians.

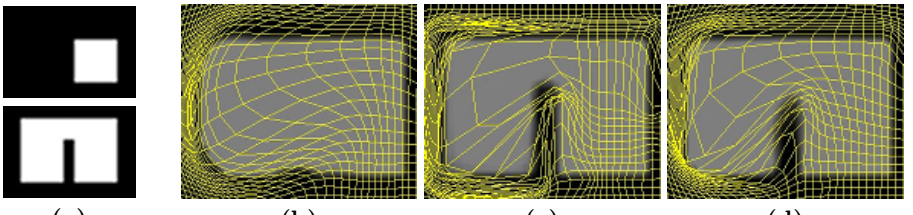


Fig. 1. Large diffeomorphic deformation of a 2D binary image. **(a)** Source image I_S and target image I_T and deformations using: **(b)** a large kernel, **(c)** a small kernel and **(d)** the sum of these kernels.

As expected, at a large scale the square is registered on the rectangle with a small perturbation while at a small scale the registration fully resolves the shape differences. Using the sum of these kernels, the deformation appears relatively similar to the one obtained using only the small kernel, even though the matching quality is slightly lower. However, as shown in the deformation grids of Fig. 1 the multi-kernel approach generates deformations that are clearly smoother and more intuitive than those obtained using the small kernel for similar final deformation. This is due to the simultaneous consideration of two scales all along the flow of deformation. Hence, in order to match objects with several scales of interest, our multi-kernel technique appears efficient since the registration of small and large scales is treated simultaneously with appropriate kernels. More discussions are given in [11].

3.2 Application to 3D Brain MR Images

To assess the ability of the proposed method in dealing with medical images, we applied it to the discrimination of MR brain images of patients with Alzheimer's disease (AD) and healthy age-matched controls. For each subject group, we

downloaded 30 pairs of images taken at baseline and after 24 months follow-up from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study [12]. All T1-weighted 1.5T MR-images were pre-processed using the standard ADNI pipeline. We aligned all images with the MNI152 brain template using affine registration before extracting a region of interest of 128^3 voxels around the hippocampus. By using this strategy instead of more accurate intra-subject pre-registrations, the baseline and follow-up images present variations due to the atrophy of the tissues and due to the inaccurate pre-registration. Therefore, comparing them at large and small scales simultaneously and quantifying the differences at small scale only is pertinent. For each subject, we then registered the baseline and follow-up MR images using several strategies: **(LK)** Using a large kernel of characteristic size $\sigma_1 = 25$mm which is obviously higher than the deformations we want to measure, **(SK)** using a small kernel of characteristic size $\sigma_2 = 1.5$mm closer to the apparent transformations, **(SGK)** using the sum of these Gaussian kernels with apparent weights $a'_2 = 2a'_1$ and means of g_1 and g_2 in each group, and finally **(LSK)** using the small kernel σ_1 to register the deformed baseline images resulting from the deformation (LK), with the follow-up image. This last strategy corresponds to a coarse to fine approach.

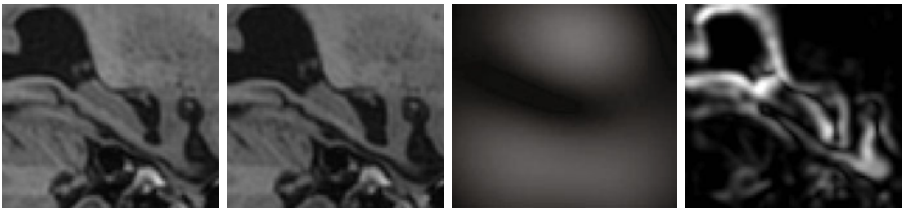


Fig. 2. (Left) Baseline and follow-up MR images of the hippocampus out of an AD subject. (Right) AOD using the sum of kernel. The contributions SGK_1 and SGK_2 of the large and small kernels is separated here.

We compared the temporal changes in hippocampus using several descriptors by automatically segmenting the hippocampus in all baseline images and transforming these volumes using the flow of deformations resulting from the registration of the grey level images. We measured, for each subject, the hippocampal atrophy using: $A = (V_s - V_t)/V_s$, where V_s and V_t are the hippocampal volumes in the segmented baseline image and after deformation to the second time point respectively. We also measured the amplitude of the deformations (AOD) from each point x of the hippocampal surface and separated the contributions of the large (SGK_1) and the small (SGK_2) kernels for (SGK) using the technique described in subsection 2.3. Here the AOD have been non-dimensionalized by the cubic root of V_s to remove the bias due to inter-subject variations of the initial hippocampal volume. For comparison purpose, we also computed from the same points of the surface the determinant of the Jacobian (det(J)) of the final deformations ϕ_1. Group-dependent average values and P-values of

non-parametric Mann-Whitney tests between the groups, which measure the discriminative power of each descriptor, are presented in table 1.

The global measure of the atrophy is similar to the measures reported in the literature [12] for (LSK) and (SGK) and turns out to be the most discriminant descriptor. The local measures of displacement AOD and det(J), averaged on the surfaces, appear to have a similar discriminant power, with slightly better results for AOD. In AOD: **(LK)** The differences between the groups reflect the variations at large scale but the influence of the small scales is not completely negligible. **(SK)** The discrimination is surprisingly good but biased by the deformations at large scale. **(LSK)** The discrimination is good. **(SGK)** The discriminative power is in between (LK) and (LSK) if the deformations (SGK_1) and (SGK_2) are not distinguished. However, when separated, the large structures are registered with few influence of the small ones and vice-versa so the discrimination is better using (SGK_2) than (LSK). Note that by using $a_2' = a_1'$ the results are similar with a slightly smaller discriminative power of SGK_2. Similarly, in LSK the image is resampled after the registration at large scale. The results would be improved by integrating the first deformation in ϕ_0 instead.

Table 1. Average values of atrophy, AOD and determinant of the Jacobian in the hippocampal, observed in the AD and healthy (H) groups. P-values of non-parametric Mann-Whitney tests measure the discriminative power of each descriptor. The lower the P-value, the clearer the distinction.

	% Atrophy			Average AOD			Average det(J)		
	mean AD	mean H	P-val.	mean AD	mean H	P-val.	mean AD	mean H	P-val.
LK	1.9	1.4	0.68	0.67	0.56	0.054	0.99	0.99	0.95
SK	8.2	1.4	1.3e-6	0.57	0.30	2.2e-5	0.96	0.99	2.1e-4
LSK	9.6	2.8	1.9e-6	0.51	0.29	2.8e-4	0.96	0.98	0.06
SGK	8.1	2.0	3.6e-7	0.72	0.52	1.7e-3	0.96	0.99	6.9e-5
SGK_1	-	-	-	0.47	0.46	0.49	-	-	-
SGK_2	-	-	-	0.41	0.20	2.4e-6	-	-	-

4 Conclusion

We have presented an approach of fine and coarse registration in the context of LDDMM for 3D medical images. This approach is of particular interest since the images often contain anatomical structures that exhibit variations at several scales of interest that cannot be properly measured using a simple or a coarse to fine LDDMM registration. The approach we developed makes use of the sum of several Gaussians to define the kernel underlying the deformations. Our tests have shown that this method estimates natural-looking deformations when registering images presenting variations at two different scales while keeping diffeomorphic properties. More importantly, it appears that our fine and coarse approach may have better abilities than standard coarse to fine approaches to

discriminate two groups of images by measuring the amplitudes of the deformations generated at a scale of interest. Note, that these scale-dependent amplitudes of deformations are local, so localized measures of shape evolution can be performed. Future work will consist in the validation of a multi-resolution approach and the assessment of the geodesic properties of the estimated deformations. A deeper characterization of the apparent weights is also an important perspective of this work. More experiments and applications of our technique will also be carried out on MR cerebral images as well as on CT cardiac images.

References

1. 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)
2. Ashburner, J.: A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113 (2007)
3. Avants, B.B., Epstein, C.L., Grossman, M., Gee, J.C.: Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis 12, 26–41 (2008)
4. Beg, F.M., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. International Journal of Computer Vision 61(2), 139–157 (2005)
5. Crum, W., Tanner, C., Hawkes, D.: Anisotropic multi-scale fluid registration: evaluation in magnetic resonance breast imaging. Physics in Medicine and Biology 50(21), 5153–5174 (2005)
6. Dupuis, P., Grenander, U., Miller, M.I.: Variational problems on flows of diffeomorphisms for image matching. Q. Appl. Math. LVI(3), 587–600 (1998)
7. Haber, E., Modersitzki, J.: Cofir: coarse and fine image registration. In: SIAM Real-Time PDE-Constrained Optimization, pp. 37–49 (2007)
8. Hernandez, M., Bossa, M.N.,Olmos, S.: Registration of anatomical images using paths of diffeomorphisms parameterized with stationary vector field flows. Int. J. Comput. Vision 85(3), 291–306 (2009)
9. Lorenzen, P., Prastawa, M., Davis, B., Gerig, G., Bullitt, E., Joshi, S.: Multi-modal image set registration and atlas formation. Med. Image Anal. 10(3), 440–451 (2006)
10. Miller, M., Younes, L.: Group actions, homeomorphisms, and matching: A general framework. International Journal of Computer Vision 41(1-2), 61–84 (2001)
11. Risser, L., Vialard, F., Murgasova, M., Holm, D., Rueckert, D.: Large diffeomorphic registration using fine and coarse strategies. application to the brain growth characterization. In: International Workshop on Biomedical Image Registration -WBIR, pp. 186–197 (2010)
12. Schuff, N., Woerner, N., Boreta, L., Kornfield, T., Shaw, L.M., Trojanowski, J.Q., Thompson, P.M., Jack, J.C.R., Weiner, M.W.: Disease Neuroimaging Initiative: MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 132(4), 1067–1077 (2009)
13. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage 45(1), S61–S72 (2009)

Registration of Longitudinal Image Sequences with Implicit Template and Spatial-Temporal Heuristics

Guorong Wu[1], Qian Wang[1,2], Hongjun Jia[1], and Dinggang Shen[1]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
{grwu,jiahj,dgshen}@med.unc.edu
[2] Department of Computer Science, University of North Carolina at Chapel Hill
qianwang@cs.unc.edu

Abstract. Accurate measurement of longitudinal changes of anatomical structure is important and challenging in many clinical studies. Also, for identification of disease-affected regions due to the brain disease, it is extremely necessary to register a population data to the common space simultaneously. In this paper, we propose a new method for simultaneous longitudinal and groupwise registration of a set of longitudinal data acquired from multiple subjects. Our goal is to 1) consistently measure the longitudinal changes from a sequence of longitudinal data acquired from the same subject; and 2) jointly align all image data (acquired from all time points of all subjects) to a hidden common space. To achieve these two goals, we *first* introduce a set of temporal fiber bundles to explore the spatial-temporal behavior of anatomical changes in each longitudinal data of the same subject. *Then*, a probabilistic model is built upon the hidden state of spatial smoothness and temporal continuity on the fibers. *Finally*, the transformation fields that connect each time-point image of each subject to the common space are simultaneously estimated by the expectation maximization (EM) approach, via the maximum a posterior (MAP) estimation of probabilistic models. Promising results are obtained to quantitatively measure the longitudinal changes of hippocampus volume, indicating better performance of our method than the conventional pairwise methods.

1 Introduction

Longitudinal study of human brains is important to reveal subtle structural and functional changes due to the brain disease [1, 2]. Although many image registration algorithms have been developed, most of them, regardless of feature-based or intensity-based, register the serial images independently when applied to longitudinal data. However, the independent registration implies the independent measurement of brain changes in serial images, and could lead to inconsistent longitudinal measurement especially for the small structures with tiny annual changes (e.g., atrophy in hippocampus due to aging [3]).

Several methods have been reported to measure the temporal anatomical changes. Shen and Davatzikos [3] proposed the 4D-HAMMER registration to cater for the longitudinal application. However, one limitation of their method is to build a 4D template by repeating one specific 3D image as templates for different time-points, which may introduce bias in longitudinal data analysis. Recently, Durrleman *et al.* [4] presented a spatial-temporal atlas estimation method to analyze the variability of the

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 618–625, 2010.

longitudinal shapes. This method uses the regression model to infer the shape evolution, and then jointly estimates the deformations as well as the dynamics of temporal changes. However, this method also needs to explicitly specify a template in advance.

In this paper, we propose a novel longitudinal registration method for image sequences acquired at multiple time-points of multiple subjects. Our method integrates both advantages of groupwise registration and longitudinal registration, thus achieving the spatial-temporal consistency for the registration results. Specifically, the transformation fields of all images, regardless of subjects and acquisition times, are *first* simultaneously estimated towards a hidden common space. *Second*, the spatial-temporal consistency of deformations within serial images of each subject is regulated by a set of temporal fiber bundles [5] to convey the spatial dependence and temporal continuity. *Finally*, these two steps are naturally integrated by a probabilistic model on fiber bundles and further solved via an EM framework. Promising results of the proposed method are obtained on both simulated and real longitudinal brain data with hippocampus atrophy, indicating better performance of our method than the conventional pairwise registration methods.

2 Methods

Given a set of longitudinal image sequences, $\boldsymbol{I} = \{I_{s,t} | s = 1, \dots, N, t = 1, \dots, T_s\}$ where $I_{s,t}$ represents the image acquired from subject s at time point t, our first goal is to align them to a common space following the transformation field $f_{s,t}$ defined from the common space to the individual image space of $I_{s,t}$. We define the mean shape of all aligned brain boundaries as $Z = \{z_j | j = 1, \dots, K\}$, where z_j denotes the position of the j-th point of the mean shape in the common space. Then, for each z_j, its transformed positions in the space of subject s, denoted as $\Phi_{s,j} = \{f_{s,t}(z_j) | t = 1, \dots, T_s\}$, can be regarded as a temporal fiber within the subject s, which facilitates the regulation of the temporal continuity for the longitudinal data. Fig. 1 illustrates the main idea of our registration method. As shown in Fig. 1(a), all corresponding points in the longitudinal sequences of different subjects (i.e., colored dots) are linked to the common space (i.e., red star) based on their respective deformation fields $f_{s,t}$, and also their longitudinal changes within each subject (i.e., dash curves) are captured by the temporal fiber bundles $\{\Phi_{s,j}\}$. Fig. 1(b) shows more details on registering a set of longitudinal data. Specifically, the mean shape Z is given in the top, with all its instances $f_{s,t}(Z)$ shown in the middle, where all images of each longitudinal data are connected by the embedding temporal fibers $\{\Phi_{s,j}\}$. To establish the point correspondences w.r.t. the common space, only the most distinctive voxels, called as driving voxels $X_{s,t}$, are used for to drive the transformation of each image $I_{s,t}$. Examples of driving voxels are given in the bottom of Fig. 1(b) as red points overlaid on brain slices. Note that each transformation field $f_{s,t}$ is steered by the correspondences determined between $X_{s,t}$ and $f_{s,t}(Z)$, with the estimated transformation fields of the same subject regularized along each temporal fiber $\Phi_{s,j}$, to achieve the temporal continuity. The hidden states of spatial-temporal characteristics on fiber bundles will be described by a probabilistic model introduced in Section 2.1. We will present a solution to this probabilistic model in Section 2.2, where the estimation of transformation fields $f_{s,t}$ is regarded as the MAP estimation of model parameters.

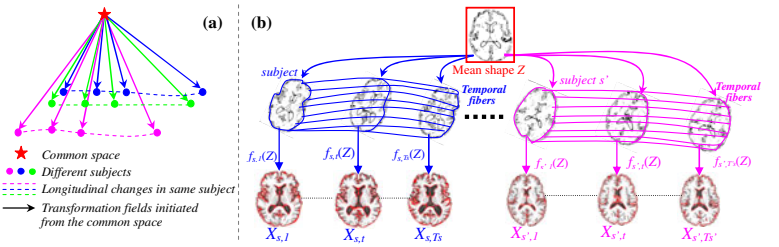


Fig. 1. The schematic illustration of the proposed method for simultaneous longitudinal and groupwise registration. In (a), not only all subjects are jointly warped onto the common space, but also the longitudinal changes are consistently measured by fibers in longitudinal data of each subject. In (b), an overview of our method is provided. For each image, the driving voxels $X_{s,t}$ (shown in red in the bottom) will be used to steer the estimation of transformation field $f_{s,t}$, by establishing their correspondences with the mean shape in the common space (as shown in the top). Meanwhile, the temporal consistency within each longitudinal data can be regulated along the temporal fiber bundles, as displayed by lines in the middle of (b).

2.1 Probabilistic Model for Registration of Longitudinal Data Sets

Our registration aims to simultaneously minimize the variability of serial images of all subjects, while enforcing the temporal continuity within the longitudinal data of each subject. In general, the transformation field $f_{s,t}$ (defined in the common space) is used to pull each data $I_{s,t}$ from its own space to the common space. The bending energy is used as a smoothness constraint to regularize each $f_{s,t}$, since thin-plate splines (TPS) [6] is efficient to give the optimal $f_{s,t}$ with minimal bending energy. In the following, we will use the operator $L_S(f_{s,t})$ to denote the bending energy of each transformation field $f_{s,t}$ in the spatial domain.

Gaussian Mixture Model on the Driving Voxels: Attribute vector has been widely used as a morphological signature to guide image registration. In this paper, we follow the method in [7] to calculate an attribute vector $\vec{a}(x)$ at each location $x \in \Re^3$ in each subject image $I_{s,t}$. Based on attribute vectors, the set of driving voxels in $I_{s,t}$, denoted as $X_{s,t} = \{X_{s,t}^i | i = 1, \ldots, M_{s,t}\}$ where $M_{s,t}$ is the total number of the driving voxels in $I_{s,t}$, can be selected for steering the registration [7]. Here, we regard each $X_{s,t}^i$ as the random variable drawn from a Gaussian mixture model (GMM) whose centers are the deformed point set $f_{s,t}(Z)$. Accordingly, we introduce a hidden variable $m_{s,t}^i \in \{1, \ldots, K\}$ (where K is the number of temporal fibers) to indicate the spatial assignment of a specific GMM center in generating $X_{s,t}^i$. The complete density function of $X_{s,t}^i$ can be given by:

$$P(X_{s,t}^i | m_{s,t}^i = j, f_{s,t}, Z_j) = N(X_{s,t}^i; f_{s,t}(Z_j), \Sigma_{s,t}^j) \propto \exp\left\{-\frac{\|X_{s,t}^i - f_{s,t}(Z_j)\|^2}{\sigma^2}\right\}, \quad (1)$$

where $\Sigma_{s,t}^j$ is the covariance matrix of the normal distribution $N(X; \mu, \Sigma)$. For sake of simplicity, the isotropic covariance matrix $\Sigma_{s,t}^j = \sigma^2 \cdot I_{K \times K}$ is assumed.

Given the spatial assignment $m_{s,t}^i$, the attribute vector on $X_{s,t}^i$, i.e., $\vec{a}(X_{s,t}^i)$, can be also viewed as a random variable generated from another GMM whose centers are the

respective attribute vectors on all other images, denoted as $\vec{a}\left(f_{s',t'}(Z_j)\right)$ $(s' \neq s, t' \neq t)$. We introduce a subject assignment $n_{s,t}^i \in \{(s',t')\}$ to denote the index of image data for reference. Similarly, the conditional density function of $\vec{a}\left(X_{s,t}^i\right)$ is defined as:

$$\begin{aligned} &P\left(\vec{a}\left(X_{s,t}^i\right)|X_{s,t}^i, m_{s,t}^i = j, n_{s,t}^i = (s',t'), f_{s,t}, f_{s',t'}, Z_j\right) \\ &= N\left(\vec{a}\left(X_{s,t}^i\right); \vec{a}\left(f_{s',t'}(Z_j)\right), \Sigma_{s,t}^j\right) \propto \exp\left\{-\frac{\left\|\vec{a}\left(X_{s,t}^i\right) - \vec{a}\left(f_{s',t'}(Z_j)\right)\right\|^2}{\tau^2}\right\}, \end{aligned} \tag{2}$$

where we assume again the isotropic covariance matrix.

By denoting the occupancy probabilities of $m_{s,t}^i$ and $n_{s,t}^i$ as $P\left(m_{s,t}^i = j\right) = \pi_{s,t}^i$ and $P\left(n_{s,t}^i = (s',t')\right) = \omega_{s,t}^i$, respectively, the posterior probability of $\left(X_{s,t}^i, \vec{a}\left(X_{s,t}^i\right)\right)$ given the transformation field $f_{s,t}$ and the mean shape Z can be given as:

$$\begin{aligned} P\left(X_{s,t}^i, \vec{a}\left(X_{s,t}^i\right)|f_{s,t}, Z_j\right) &= \sum_{s=1}^{N}\sum_{j=1}^{K} \pi_{s,t}^i \cdot \omega_{s,t}^i \cdot \exp\left(-\xi_{s,t}^i\left(\pi_{s,t}^i, \omega_{s,t}^i\right)\right) \\ \xi_{s,t}^i\left(\pi_{s,t}^i, \omega_{s,t}^i\right) &= \frac{\left\|X_{s,t}^i - f_{s,t}(Z_j)\right\|^2}{\sigma^2} + \frac{\left\|\vec{a}\left(X_{s,t}^i\right) - \vec{a}\left(f_{s',t'}(Z_j)\right)\right\|^2}{\tau^2}. \end{aligned} \tag{3}$$

Spatial-Temporal Heuristics on Fiber Bundles: For each subject $\boldsymbol{s}$, the temporal fiber bundles capture the longitudinal consistency of the estimated deformations, as shown in Fig. 1. In other words, we can take advantage of the temporal fiber bundles, designated by $(\boldsymbol{f_{s,t}}, \boldsymbol{t = 1, \dots, T_s})$ from the common space, to regulate the temporal smoothness from $\boldsymbol{f_{s,1}(Z_j)}$ to $\boldsymbol{f_{s,T_s}(Z_j)}$ along the $\boldsymbol{j}$-th fiber. A traditional kernel-based approach can be used here to assure the temporal continuity. In particular, the position of $\boldsymbol{f_{s,t}(Z_j)}$ after kernel smoothing is defined by:

$$\hat{f}_{s,t}(Z_j) = \sum\nolimits_{\tau=1}^{T_s} G\left(\frac{\tau - t}{w}\right) \cdot f_{s,\tau}(Z_j) \Big/ \sum\nolimits_{\tau=1}^{T_s} G\left(\frac{\tau - t}{w}\right), \tag{4}$$

where G is a Gaussian kernel and w the kernel width. The principle of Gaussian kernel smoothing leads to the minimization of the deviation between $f_{s,t}(Z_j)$ and $\hat{f}_{s,t}(Z_j)$. We use the operator $L_T\left(f_{s,t}(Z_j)\right) = \left\|\hat{f}_{s,t}(Z_j) - f_{s,t}(Z_j)\right\|^2$ to measure the residual energy after performing temporal smoothing on each position of a fiber, which accounts for the smoothness in temporal domain. It is worth noting that more sophisticated kernel-based regression methods are also applicable here.

Unified Probabilistic Model for Registration of Longitudinal Data Sets: Considering $h = \left(\{Z_j\}, \{f_{s,t}\}, \{m_{s,t}^i\}, \{n_{s,t}^i\}\right)$ as the model parameters and $v = \left(\{X_{s,t}^i\}, \{\vec{a}(X_{s,t}^i)\}\right)$ as the observation, the joint probability $P(h, v)$ is given by:

$$\begin{aligned} P(h,v) &\propto \left[\prod_{j=1}^{K} P(Z_j)\right] \cdot \left[\prod_{s=1}^{N} P(f_{s,t}) \cdot \prod_{j=1}^{K}\prod_{t=1}^{T_s} P\left(f_{s,t}(Z_j)\right)\right] \cdot \left[\prod_{i=1}^{M_{s,t}} P\left(X_{s,t}^i, \vec{a}\left(X_{s,t}^i\right)|f_{s,t}, Z_j\right)\right], \\ P(f_{s,t}) &= e^{-\frac{\lambda_1}{2}\left\|L_S(f_{s,t})\right\|^2}, P\left(f_{s,t}(Z_j)\right) = e^{-\frac{\lambda_2}{2}\left\|L_T\left(f_{s,t}(Z_j)\right)\right\|^2}, \end{aligned} \tag{5}$$

where $P\left(Z_j\right) = \frac{1}{K}$ due to no prior knowledge on Z_j.

Hereafter, our registration problem turns to infer the model parameters h that best interpret the observation v, i.e., $\hat{h} = \arg\max_h P(h, v)$.

2.2 Solutions to Our Longitudinal Registration

Finding the best model parameters $\hat{h}$ by MAP estimation on $P(h, v)$ is intractable since many parameters h might explain the observation v very well. "Free energy with annealing" [8] is one of the variational Bayes inference approach to solve this problem by minimizing the distribution distance between $p(h|v)$ and $p(h, v)$, which bounds for maximizing the likelihood of the observation $P(v)$. However, here the posteriori distributions of some variables ($f_{s,t}$, $\hat{f}_{s,t}$, and Z) are hard to compute. Instead, we perform the point estimation on $f_{s,t}$, $\hat{f}_{s,t}$, and Z by maximizing their expectation w.r.t. the maintained posteriori distributions of $\{m_{s,t}^i\}$ and $\{n_{s,t}^i\}$, which falls to the EM principle. By omitting some constant terms and taking logarithm to Eq. 5, we obtain the free energy for our registration algorithm as:

$$E(\pi, \omega, f, \hat{f}, Z) \propto \sum_{s=1}^{N}\sum_{t=1}^{T_s}\sum_{j=1}^{K}\sum_{i=1}^{M_{s,t}}\left[\pi_{s,t}^i \cdot \omega_{s,t}^i \cdot \xi_{s,t}^i(\pi_{s,t}^i, \omega_{s,t}^i) + \left(r_1 \cdot \pi_{s,t}^i \cdot \log(\pi_{s,t}^i) + r_2 \cdot \omega_{s,t}^i \cdot \log(\omega_{s,t}^i)\right)\right] + \sum_{s=1}^{N}\frac{\lambda_1}{2}L_s(f_{s,t}) + \frac{\lambda_2}{2}\sum_{s=1}^{N}\sum_{j=1}^{K}\sum_{t=1}^{T_s}L_T\left(f_{s,t}(Z_j)\right), \tag{6}$$

where r_1 and r_2 are the "temperatures" for spatial assignment and subject assignment, respectively, in annealing scenario. It is worth noting that r_1 and r_2 will gradually decrease to encourage the assignment changing from fuzzy to strictly binary.

The spatial assignment $\pi_{s,t}^i$ and subject assignment $\omega_{s,t}^i$ can be obtained by minimizing E w.r.t. $\pi_{s,t}^i$ and $\omega_{s,t}^i$, respectively:

$$\hat{\pi}_{s,t}^i = \exp\left\{-\frac{\omega_{s,t}^i}{r_1}\cdot\left(\xi_{s,t}^i(\pi_{s,t}^i, \omega_{s,t}^i)\right)\right\}, \qquad \pi_{s,t}^i \leftarrow \hat{\pi}_{s,t}^i \Big/ \sum_{j=1}^{K}\hat{\pi}_{s,t}^i \tag{7}$$

$$\hat{\omega}_{s,t}^i = \exp\left\{-\frac{\pi_{s,t}^i}{r_2}\cdot\left(\xi_{s,t}^i(\pi_{s,t}^i, \omega_{s,t}^i)\right)\right\}, \qquad \omega_{s,t}^i \leftarrow \hat{\omega}_{s,t}^i \Big/ \sum_{s=1}^{N}\hat{\omega}_{s,t}^i \tag{8}$$

Next, Eq. 6 can be optimized by decoupling it into two sub-problems, i.e., (**SP$_1$**) tracking the longitudinal changes along the fiber bundles; and (**SP$_2$**) simultaneously estimating the transformation fields $f_{s,t}$ and updating the mean shape Z.

SP$_1$: Longitudinal Tracking along Fiber Bundles: Once the mean shape Z and transformation $f_{s,t}$ of each image have been determined, temporal continuity is preserved by tracking the composite variables from $f_{s,1}(Z_j)$ to $f_{s,T_S}(Z_j)$ in each temporal fiber Z_j. Here, kernel-based smoothing is performed along each fiber of individual subject, to enforce the temporal continuity by Eq. 4 and minimizes the energy function E in Eq. 6 w.r.t. $f_{s,t}(Z_j)$.

SP$_2$: Groupwise Registration: This step aims to refine $f_{s,t}$ and Z by minimizing the energy function E with fixed π, ω, and $\hat{f}_{s,t}$. This can be considered as a groupwise registration problem, which is solved in three iterative steps. First, the transformation $f_{s,t}(Z_j)$ w.r.t. to all driving voxels $X_{x,t}^i$ in $I_{s,t}$ is updated as a weighted mean location according to the observations of $\pi_{s,t}^i$ and $\omega_{s,t}^i$, i.e., $f_{s,t}(Z_j) = \sum_{i=1}^{M_{s,t}}\pi_{s,t}^i \cdot \omega_{s,t}^i \cdot X_{s,t}^i$. Second, by considering $\{Z_j\}$ as the source point set and $\{f_{s,t}(Z_j)\}$ as the target point set, TPS interpolates the dense transformation field $f_{s,t}$ for each $I_{s,t}$. Finally, by

minimizing E with respect to Z_j, the mean shape Z is updated as $\hat{Z}_j = \sum_{s=1}^{N}\sum_{t=1}^{T_s}\hat{\omega}_{s,t}^{i}\sum_{i=1}^{M_{s,t}}\hat{\pi}_{s,t}^{i}f_{s,t}^{-1}\left(X_{s,t}^{i}\right)$, where $f_{s,t}^{-1}$ is the inverse of the transformation field $f_{s,t}$.

3 Experiments

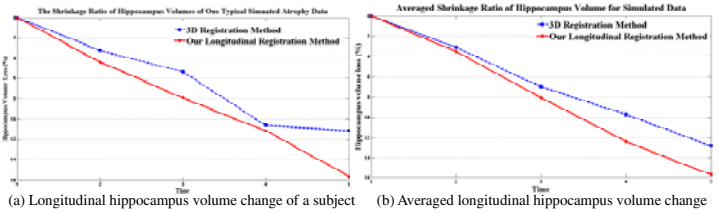
The ultimate goal of our longitudinal registration algorithm is to consistently measure the subtle anatomical changes to facilitate the diagnosis (or early prediction) of diseases, i.e., brain dementia. To evaluate the performance of our method, two sets of data (one with simulated atrophy in hippocampus, and another with manually labeled hippocampi in year 1 and year 5 of 9 elder brains) are used for demonstration in this section. For comparison, we have also used a 3D pairwise registration algorithm [9] to independently measure the longitudinal changes of hippocampi for these data.

Experiments on Simulated Atrophy in Hippocampus: Three subjects with manually labeled hippocampus regions at year 1 are used as the baseline data ($t = 1$) to simulate atrophy on their hippocampi. Then, we yield the next time-point image ($t = 2$) by simulating ~5% hippocampal shrinkage. Repeating this procedure, we finally obtain 3 sets of longitudinal data with hippocampal volume shrinking at an annual rate of ~5% in the five years. Considering the numerical error in simulation, the total simulated atrophy is 15.86% within the five years.

Our algorithm can simultaneously estimate the transformation fields of these 5×3 images from their own spaces to the hidden common space, and keep spatial-temporal continuity on the estimated transformation fields for each subject. Fig. 2(a) shows the curves of the estimated hippocampus volume loss by a 3D pairwise registration method and our proposed method in blue and red curves, respectively. It is obvious that our method outperforms the 3D pairwise registration method, with more consistent measurement of longitudinal changes. The averaged hippocampus volume loss estimated by the two methods is also displayed in Fig. 2(b). The final estimated hippocampus volume loss is 12.8% by the 3D pairwise registration method and 15.73% by our method, which shows the capability of our method in more accurate measurement of anatomical changes.

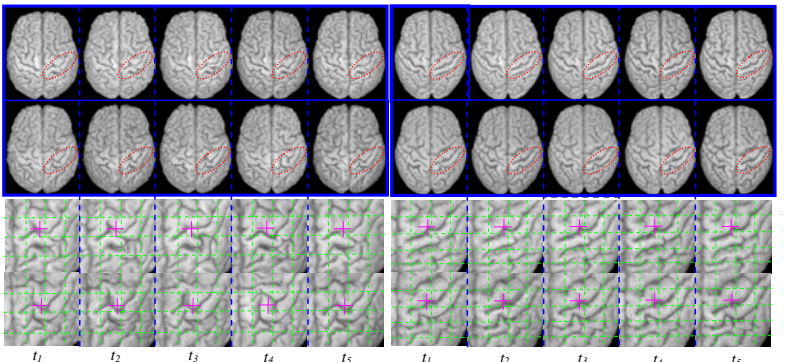
Experiments on Real Elderly Brains: Nine elderly subjects, each with annual MR scans in 5 years, are evaluated in this experiment. For comparison, we independently warp these 9×5 subjects to a selected template by the 3D pairwise registration method. The registration results of all five time-points of the two subjects are shown in the top two rows of Fig. 3(a). We also perform our longitudinal registration method to simultaneously align these 45 subjects to the hidden common space without specifying the template. The registration results of the same two subjects across 5 years are displayed in the top two rows of Fig. 3(b). It can be observed that our method achieves more consistent alignment in global brain shapes. We further enlarge the right pre-/post-central gyrus area, as shown in two bottom rows. As indicated by the pink crosses, the registration results along the time axis are not continuous for the 3D method, while the longitudinal registration results of our method are more consistent, due to the introduction of the spatial-temporal heuristics. Noting that the percentage

of hippocampal shrinkage detected from year 1 to year 5 is 3.84% by 3D method and 5.38% by our method, compared to 5.65% by two experienced raters.



(a) Longitudinal hippocampus volume change of a subject (b) Averaged longitudinal hippocampus volume change

Fig. 2. Experimental results on measuring simulated hippocampus volume loss in the 5 years. (a) shows the hippocampus volume loss for a typical subject in the 5 years and (b) shows the averaged hippocampus volume change in the 5 years. The total simulated atrophy in the 5 years is 15.86%. The estimated atrophy is 15.73% by our method and 12.8% by the 3D method, indicating more accurate results achieved by our method.



(a) By 3D registration method (b) By our proposed method

Fig. 3. Registration results on two typical subjects (with 5 scans) by 3D pairwise registration method (a) and our proposed method (b). The top two rows show the 3D rendering of two aligned subjects in 5 time points. Visually, our method achieves more consistent results in the global brain shapes (particularly for the regions indicated by the red ellipses). The bottom two rows show an enlarged region at right pre-/post-central gyri area. As indicated by pink crosses, our method outperforms the pairwise registration method in temporal continuity.

The registration performance can be further evaluated by computing the overlap ratio on brain tissues between each aligned image and the reference image. In our algorithm, since no explicit template is used, the reference image is voted from all aligned images by setting the tissue at each voxel as the majority of tissue labels from all aligned images. First, by considering all 45 aligned images together, the average overlap ratio of white matter (WM), gray matter (GM), ventricle (VN) are 67.2%, 59.2%, 79.0% by the 3D method, while 72.6%, 63.9%, and 82.1% by our method. Second, by considering all 5 aligned scans of each subject as a separate group, the mean and standard deviation of overlap ratios from 9 subjects are ($79.7\% \pm 2.5\%$), ($71.3\% \pm 2.4\%$), ($86.3\% \pm 1.7\%$) by the 3D method on WM, GM, and VN,

respectively, compared to $(82.1\% \pm 1.4\%)$, $(74.5\% \pm 2.1\%)$, $(86.6\% \pm 1.3\%)$ by our method. Obviously, these two sets of results consistently demonstrate that our method outperforms the 3D method in accurately aligning all 45 images to the common space (first experiment), as well as better preserving the temporal continuity in each subject (second experiment).

4 Conclusion

We have presented a novel method for longitudinal registration of serial images, by taking advantages of both spatial-temporal heuristics and implicit template in registering a group of longitudinal data. Specifically, the spatial-temporal continuity is achieved by enforcing smoothness constraint along fiber bundles in each subject. Further, we simultaneously estimate all transformation fields which link all images to the hidden common space without specifying any template, thus avoid introduction of any bias to subsequent data analysis.

In future, we will investigate more powerful kernel regression methods to regulate the temporal smoothness along the temporal fiber bundles. Also, we will test our longitudinal registration method in clinical applications, e.g., measuring longitudinal brains change in MCI study.

References

1. Cardenas, V.A., Du, A.T., Hardin, D., Ezekiel, F., Weber, P., Jagust, W.J., Chui, H.C., Schuff, N., Weiner, M.W.: Comparison of methods for measuring longitudinal brain change in cognitive impairment and dementia. Neurobiology of Aging. 24(4), 537–544 (2003)
2. Toga, A.W., Thompson, P.M.: Temporal dynamics of brain anatomy. Annu. Rev. Biomed. Eng. 5, 119–145 (2003)
3. Shen, D., Davatzikos, C.: Measuring Temporal Morphological Changes Robustly in Brain MR Images Via 4-Dimensional Template Warping. NeuroImage 21(4), 1508–1517 (2004)
4. Durrleman, S., Pennec, X., Trouve, A., Gerig, G., Ayache, N.: Spatiotemporal Atlas Estimation for Developmental Delay Detection in Longitudinal Datasets. In: Yang, G., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009, Part I. LNCS, vol. 5761, pp. 297–304. Springer, Heidelberg (2009)
5. Di, H., Tao, L., Xu, G.: A Mixture of Transformed Hidden Markov Models for Elastic Motion Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 31(10), 1817–1830 (2009)
6. Bookstein, F.L.: Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. IEEE Transactions on Pattern Analysis and Machine Intelligence 11(6), 567–585 (1989)
7. Shen, D., Davatzikos, C.: HAMMER: Hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging 21(11), 1421–1439 (2002)
8. Frey, B.J., Jojic, N.: A Comparison of Algorithms for Inference and Learning in Probabilistic Graphical Models. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(9), 1392–1416 (2005)
9. Wu, G., Yap, P.T., Kim, M., Shen, D.: TPS-HAMMER: Improving HAMMER registration algorithm by soft correspondence matching and thin-plate splines based deformation interpolation. NeuroImage 49(3), 2225–2233 (2010)

3D Ultrasound to Stereoscopic Camera Registration through an Air-Tissue Boundary

Michael C. Yip[1], Troy K. Adebar[1], Robert N. Rohling[1], Septimiu E. Salcudean[1], and Christopher Y. Nguan[2]

[1] University of British Columbia, British Columbia V6T1Z4 Canada
[2] Vancouver General Hospital, Vancouver, British Columbia V5Z1M9 Canada

Abstract. A novel registration method between 3D ultrasound and stereoscopic cameras is proposed based on tracking a registration tool featuring both ultrasound fiducials and optical markers. The registration tool is pressed against an air-tissue boundary where it can be seen both in ultrasound and in the camera view. By localizing the fiducials in the ultrasound volume, knowing the registration tool geometry, and tracking the tool with the cameras, a registration is found. This method eliminates the need for external tracking, requires minimal setup, and may be suitable for a range of minimally invasive surgeries. A study of the appearance of ultrasound fiducials on an air-tissue boundary is presented, and an initial assessment of the ability to localize the fiducials in ultrasound with sub-millimeter accuracy is provided. The overall accuracy of registration (1.69 ± 0.60 mm) is a noticeable improvement over other reported methods and warrants patient studies.

1 Introduction

Augmented reality in surgery often involves the superposition of medical images of a patient's anatomy onto a camera-based image of the overlying tissues. With sufficient registration accuracy, the surgeon is then able to localize the internal anatomy (subsurface features such as lesions or nerves) for improved surgical guidance. Prior work in augmented reality for surgical applications has been applied to x-ray, computed tomography (CT), magnetic resonance imaging (MRI) and ultrasound (US) [1]–[7]. This paper explores the ability to display an external three-dimensional ultrasound (3DUS) volume in a laparoscopic camera view during minimally invasive surgery.

To date, registration between an US volume and camera images has generally involved two tasks: calibrating the ultrasound volume to the pose of the US transducer, and tracking both the US transducer and the cameras in 3D space. For the first task, numerous techniques for US transducer calibration have been previously investigated; Lindseth [8] and Mercier [9] offer extensive reviews on the subject of US transducer calibration. The second task, tracking of the cameras and an US transducer, can be performed using magnetic trackers, optical trackers, or robot kinematics [5]–[7]. While these techniques have proven useful in the operating room, they still have their shortcomings: time-consuming transducer

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 626–634, 2010.

calibrations, additional tracking equipment, line-of-sight issues, modifications to the US transducers and cameras for tracking, and consequently the consumption of valuable time and space in the operating room. In addition, cumulative errors in transducer calibration and equipment tracking contribute to errors in registration that may be amplified by a lever-arm effect.

We address the above issues by introducing a new technique for registering stereoscopic cameras and 3DUS using a registration tool. Fiducials attached to the registration tool are held against the air-tissue boundary and imaged through the tissue. Their locations in the US volume are determined, and through the known geometry of the tool and the tracking of the tool by the stereoscopic cameras, a 3DUS to camera registration can be found. This provides a direct transformation from the US to the stereoscopic cameras, thus eliminating errors related to calibrating the US transducer and tracking the US transducer and the cameras. Registration of 3DUS to the cameras directly across the air-tissue boundary is the key innovation of this paper.

We envisage using the novel registration system in laparoscopic surgery, where structures being operated on can be imaged more effectively with external or endocavity US, or with laparoscopic US from a different direction than the camera view. For example, during partial nephrectomy, the kidney can be imaged with external US through the abdomen, or with laparoscopic US from one side of the kidney while the camera and laparoscopic instruments operate from the other. Similarly, in laparoscopic radical prostatectomy, endorectal US can image the prostate as in prostate brachytherapy, while the surgery proceeds with the trans-abdominal or trans-perineal approach. In both cases, registration is required to overlay the US image onto the endoscopic camera view. We propose that it be carried out with the technique described in this paper.

2 Methods

Our study had two goals: to determine the accuracy of locating US fiducials on an air-tissue boundary, and to determine the feasibility of using these fiducials to register 3DUS to stereoscopic cameras. We first examined the accuracy of localizing spherical fiducials on an air-tissue boundary in US. Air-tissue boundaries exhibit high reflection at their surfaces that may make it difficult to accurately localize fiducials. We considered five variables that could affect the accuracy of fiducial localization: (1) fiducial size, (2) lateral position in US image, (3) angle of air-tissue boundary, (4) boundary depth, and (5) stiffness of tissue.

Next, we implemented a direct closed-form registration method between 3DUS and a stereoscopic camera by localizing surface fiducials in both the 3DUS volume and the stereo camera (Figure 1a). This method is described below.

We begin by defining four coordinate systems, the stereo camera system $\{\underline{o}_0, \underline{C}_0\}$, the optical marker system $\{\underline{o}_1, \underline{C}_1\}$, the US fiducial system $\{\underline{o}_2, \underline{C}_2\}$ and the 3DUS system $\{\underline{o}_3, \underline{C}_3\}$. The transformation from $\{\underline{o}_1, \underline{C}_1\}$ to $\{\underline{o}_0, \underline{C}_0\}$, ${}^0T_1$, is found by stereo-triangulating the optical markers on the registration tool. The transformation from $\{\underline{o}_2, \underline{C}_2\}$ to $\{\underline{o}_1, \underline{C}_1\}$, ${}^1T_2$, is constant and known from

the tool geometry. The transformation from $\{\underset{\sim}{o}_3, \underline{C}_3\}$ to $\{\underset{\sim}{o}_2, \underline{C}_2\}$, ${}^2T_3$, is found by localizing three fiducials that define $\{\underset{\sim}{o}_2, \underline{C}_2\}$ in the 3DUS system $\{\underset{\sim}{o}_3, \underline{C}_3\}$.

The three fiducial locations in coordinate system $\{\underset{\sim}{o}_3, \underline{C}_3\}$, ${}^3x_0$, ${}^3x_1$ and ${}^3x_2$, define two perpendicular vectors with coordinates

$$ {}^3v_1 = {}^3x_1 - {}^3x_0 \tag{1} $$

$$ {}^3v_2 = {}^3x_2 - {}^3x_0 \ , \tag{2} $$

that can be used to define the unit vectors of frame $\underline{C}_2$ in system $\{\underset{\sim}{o}_3, \underline{C}_3\}$:

$$ {}^3i_2 = \frac{{}^3v_1}{||{}^3v_1||} \tag{3} $$

$$ {}^3k_2 = \frac{{}^3v_1 \times {}^3v_2}{||{}^3v_1 \times {}^3v_2||} \tag{4} $$

$$ {}^3j_2 = {}^3k_2 \times {}^3i_2 \ . \tag{5} $$

The origin $\underset{\sim}{o}_2$ has coordinates ${}^3x_0$ in $\{\underset{\sim}{o}_3, \underline{C}_3\}$. The homogeneous transformation from the 3DUS system $\{\underset{\sim}{o}_3, \underline{C}_3\}$ to the US fiducial system $\{\underset{\sim}{o}_2, \underline{C}_2\}$, ${}^3T_2$, is then

$$ {}^3T_2 = \begin{bmatrix} {}^3i_2 & {}^3j_2 & {}^3k_2 & {}^3x_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6} $$

and ${}^2T_3 = {}^3T_2{}^{-1}$. The overall transformation between the stereo camera system $\{\underset{\sim}{o}_0, \underline{C}_0\}$ and the 3DUS system $\{\underset{\sim}{o}_3, \underline{C}_3\}$ is then

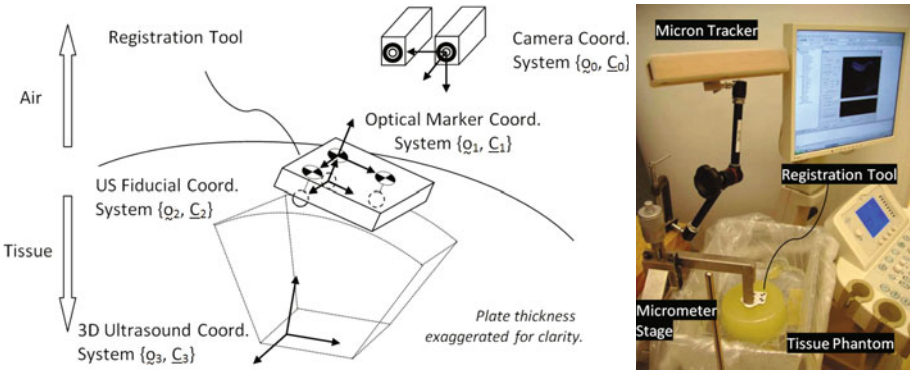
$$ {}^0T_3 = {}^0T_1 {}^1T_2 {}^2T_3 \ . \tag{7} $$

A homogenous transformation can then be constructed to register the 3DUS frame to the stereo camera frame. Lastly, with known camera parameters (focal length, image center, distortion coefficient, etc.), the registered US volume in the camera frame can be projected onto the two stereoscopic images.

2.1 Experimental Setup

Figure 1b shows the experimental setup used in this study. 3DUS volumes were captured using a Sonix RP US machine (Ultrasonix Medical Corp., Richmond, Canada) with a mechanical 3D transducer (model 4DC7-3/40). A three-axis mechanical micrometer stage was used for accurate positioning of registration tools relative to the fixed US transducer and stereo cameras.

Surface Fiducial Localization. Sets of steel spherical fiducials arranged in precisely known geometry were pressed against tissue-mimicking phantoms, imaged and localized in the 3DUS volumes. The steel plates contained three sets of fiducials spaced 10 cm apart, with each set consisting of a center fiducial and eight surrounding fiducials at a radius of 10 mm (Figure 2a). The fiducials were seated in holes cut into the plate on a water jet cutter with dimensional accuracy

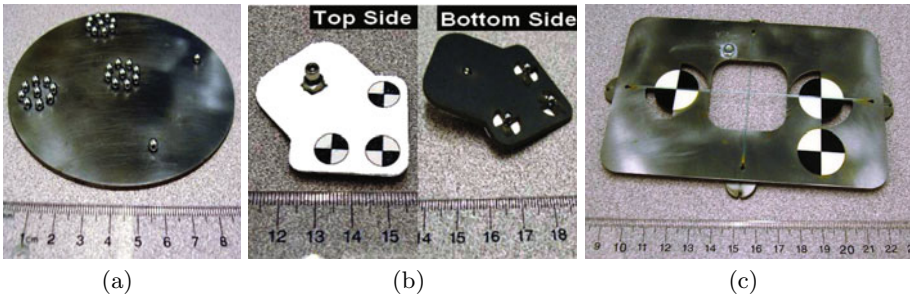


(a) (b)

Fig. 1. a) Schematic of the registration method, b) Experimental Setup

of 0.13 mm. Fiducial diameters of 2 mm, 3 mm and 4 mm were imaged through phantoms with thicknesses of 3 cm, 6 cm and 9 cm, stiffnesses of 12 kPa, 21 kPa and 56 kPa, and boundary angles of 0 degrees, 20 degrees and 40 degrees. The phantoms were made from polyvinyl chloride (PVC) using a ratio of liquid plastic to softener of 1:1 (12 kPa, low stiffness), 2:1 (21 kPa, medium stiffness), and :0 (56 kPa, high stiffness) to create phantoms that mimicked tissue properties [10]. To create fully developed speckle, one percent (mass) cellulose was added as a scattering agent.

The five independent variables evaluated were varied independently about a control case (3 mm fiducials, 6 cm depth, 21 kPa stiffness, 0 degree angle, and central location). The surface fiducial plates were pressed lightly into the PVC tissue phantoms, and imaged through the phantoms. The fiducials were then manually localized in the US volume, and the Euclidean distances between the outer fiducials and the center fiducials were compared to the known geometry to determine the accuracy of localization. For every variable level, 10 tests with 8 error measurements were performed (n = 80). The focal depth was set to the boundary depth in all tests.



(a) (b) (c)

Fig. 2. a) Fiducial localization test plate, b) Registration tool, c) Registration accuracy test tool

Registration. For the registration experiments, a Micron Tracker H3-60 optical tracking system (Claron Technology, Toronto, Canada) was used as the stereoscopic cameras. This provided a stable and accurate pre-calibrated camera system and allowed the analysis of registration accuracy to focus mainly on localizing the registration tool in US. The registration tool was also built on a steel plate cut on a waterjet cutter (Figure 2b). On the top surface are three Micron Tracker markers spaced 20 mm and 15 mm apart forming an "L" shape; on the bottom surface are three surface fiducials (3 mm) seated in holes cut directly in line with the Micron Tracker markers. Registration accuracy was measured using the test tool shown in Figure 2c. The test tool consists of a steel frame, a nylon wire cross which can be accurately localized by US in a water bath, and three Micron Tracker markers which allow the tracker to determine the location of the wire cross in the camera frame.

We first determined the homogeneous transformation relating points in the US frame to the camera frame using the registration tool. We then evaluated the accuracy of the transformation using the test tool placed in a water bath. The crosswire location in US registered into the camera frame was compared to the location of the crosswire in the camera coordinates. This was done by saving a US volume of the crosswire in a water bath, and then draining the water to track the optical markers on the test tool in the stereoscopic cameras. Registration error was defined as the Euclidean distance between the position predicted by registration and the tracked position of the crosswire in the cameras. The errors were transformed into the US frame so that they could be specified in the lateral, elevational and axial directions. To ensure accurate scales, the US volumes were generated using the correct speeds of sound for the phantoms and for water.

3 Results

The results of the fiducial localization experiments are shown in Table 1. Hypothesis testing was used to determine the statistical signficance of variables. Given two groups x and y, unpaired student t-tests determine the probability p that the null hypothesis ($\mu_x = \mu_y$) was true. A one-way analysis of variance (ANOVA) is a generalization of the student t-test for more than two groups and produces probability statistics $\{F, p\}$. For both the student t-test and ANOVA, a generally accepted probability for suggesting statistical significance is $p < 0.05$.

The ANOVA results showed that the size of the fiducial, the depth of the boundary, and the angle at which the boundary was imaged affect the accuracy of fiducial localization (ANOVA: $F_{size} = 7.34, p_{size} = 9.96E-04; F_{depth} = 15.5, p_{depth} = 1.11E-06; F_{angle} = 8.49, p_{angle} = 3.61E-04$). However, the tissue stiffness does not significantly change the accuracy of fiducial localization (ANOVA: $F_{stiffness} = 0.0414, p_{stiffness} = 0.960$). T-tests results showed that the lateral position of the fiducials on the boundary plays a significant role in the accuracy of localization (t-test: $p = 7.10E-18$).

In our registration experiment, four unique poses of the registration tool were used for a physical configuration of the camera and US transducer. The registration errors were computed at twelve different locations in the US volume. To test

Table 1. Mean, standard deviation and median of errors associated with localizing bead fiducials at air-tissue boundaries. Significantly different from control case = (*)

Variable	Value	Mean ± Std Dev. (mm)	Median (mm)	RMS Error
Fiducial Size	2 mm	0.94 ± 0.34*	0.89	1.00
	3 mm	0.82 ± 0.28	0.78	0.87
	4 mm	0.70 ± 0.20	0.67	0.73
Boundary Depth	Long (9 cm)	0.54 ± 0.18*	0.55	0.57
	Med. (6 cm)	0.82 ± 0.28	0.78	0.87
	Short (3 cm)	0.66 ± 0.20*	0.64	0.69
Tissue Stiffness	High (12kPa)	0.81 ± 0.30	0.78	0.86
	Med. (21kPa)	0.82 ± 0.28	0.78	0.87
	Low (56kPa)	0.80 ± 0.19	0.80	0.82
Boundary Angle	0°	0.82 ± 0.28	0.78	0.87
	20°	0.78 ± 0.28	0.75	0.83
	40°	1.04 ± 0.35*	0.97	1.10
Lateral Position	Center	0.82 ± 0.28*	0.78	0.87
On Boundary	Offset (10 cm)	0.60 ± 0.28*	0.59	0.66

Table 2. Mean errors (n = 12) between points in a registered 3DUS volume and its location in the stereo-camera frame

	$e_{Lateral}$ (mm)	$e_{Elevational}$ (mm)	e_{Axial} (mm)	e_{Total} (mm)
Registration 1	0.90 ± 0.44	0.77 ± 0.33	1.08 ± 0.75	1.75 ± 0.56
Registration 2	1.02 ± 0.45	0.60 ± 0.32	1.14 ± 0.99	1.83 ± 0.74
Registration 3	0.65 ± 0.43	0.76 ± 0.33	1.01 ± 0.63	1.55 ± 0.53
Registration 4	0.57 ± 0.40	0.82 ± 0.30	1.03 ± 0.79	1.60 ± 0.58
Average	0.78 ± 0.45 mm	0.74 ± 0.32 mm	1.07 ± 0.78 mm	1.69 ± 0.60 mm

the repeatability of this method, the registration was repeated four times on the same images. Table 2 shows that the average error among all the transformed points for all transformations was 1.69 mm, with a minimum error of 1.55 mm and a maximum error of 1.84 mm. The time required to perform a registration was approximately equal to the acquisition time of a 3DUS volume (2 sec).

4 Discussion

The fiducial localization tests showed that errors associated with localizing surface fiducials at an air-tissue boundary ranged from 0.54 mm to 1.04 mm. Several variables had a significant effect on accuracy. The smaller fiducials (2 mm) produced higher localization errors, suggesting that the fiducials became lost in the boundary reflection. The larger fiducials presented larger features that were easier to detect. Boundary depths farther away from the US transducer produced lower localization errors, as fiducial centers were more difficult to localize when approaching the near field [11].

Two results from the localization error analysis that have practical implications are that tissue stiffness does not significantly affect the accuracy of fiducial localization and that only large angles (e.g. 40 degrees) significantly affect the localization accuracy. Our registration method should therefore remain accurate for tissues with a wide variety of stiffnesses and shapes. The lateral location of the fiducials on the air-tissue boundary, however, was significant to the localization accuracy. The air-tissue boundary exhibited greater specular reflection near the axis of the US transducer, and thus fiducials offset laterally from the axis were less obscured by specular reflection and could be more accurately localized.

The registration experiment showed that using fiducials on an air-tissue boundary for direct registration between 3DUS and stereo cameras is feasible with an accuracy of 1.69 ± 0.60 mm. The largest errors were in the axial directions since the tail artifacts of the surface fiducials obscured the true depth at which the fiducials were located in the US volume (Figure 3). Repeated registrations on the same data and registrations using different physical locations of the registration tool all gave consistent overall and component errors, suggesting a model of the reverberation tail could improve localization and registration accuracy further. Nevertheless, based on the overall errors, our registration method is a promising alternative to using tracking equipment, where errors for similar US-to-camera registration systems are within 3.05 ± 0.75 mm [5] for magnetic tracking and 2.83 ± 0.83 mm [6] for optical tracking. It is clear that the main source of error for the new registration method is the localization of registration tool fiducials, as any localization errors would be amplified by a lever-arm effect.

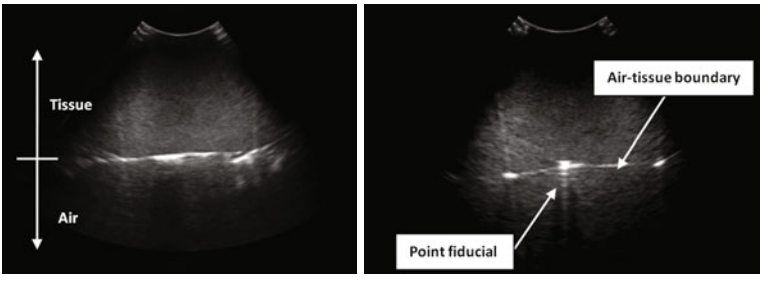


Fig. 3. Example images of an air-tissue boundary (left) and a 3 mm fiducial pressed against an air-tissue boundary (right)

The proposed registration is ideal for situations where the camera and the US transducer are fixed. However, if the US transducer or the camera is moved, a new registration can simply be acquired. Alternatively, in the case of robot-assisted surgery, the robot kinematics can be used to determine the new locations of the camera or the US transducer and maintain continuous registration to within the accuracy of robot kinematic calculations from joint angle readings.

A few practical issues with the proposed registration method should be considered. First, stereo-camera disparity plays a significant role in the accuracy

of registration. The registrations presented in this paper were performed using the Claron Micron Tracker; this represents an ideal case, as the cameras have a large disparity (12 cm) and a tracking error of $\pm$ 0.2 mm. In minimally invasive surgery, laparoscopic stereo-cameras having much smaller disparities would be used, possibly resulting in higher errors (although the cameras are imaging a much shallower depth so that the effect of disparity is lessened). This can be compensated for by maximizing the size of the registration tool, producing a well-conditioned system for computing the transformations. Such a registration tool could be designed to fold and fit through a trocar for laparoscopic surgery.

Another way to improve registration accuracy is to introduce redundancy into the registration data. Our registration tool featured only the minimum three fiducials required to extract the six degrees of freedom transformation between the US volume and the stereoscopic cameras; with more fiducials on the registration tool, averaging could be used to reduce errors. In addition, higher accuracies can be achieved by considering different poses of the registration tool in both the US and the camera frame [12].

5 Conclusions

In this study, we evaluated the accuracy of localizing fiducials pressed against an air-tissue boundary in ultrasound. We have shown that this method can be used to perform 3D ultrasound to stereo camera registration for augmented reality in surgery. This method provides a direct closed-form registration between a 3DUS volume and a stereoscopic camera view, does not require calibration of the US transducer or tracking of cameras or US transducers, and provides improved accuracies over tracking-based methods. Future work will utilize laparoscopic stereo-cameras in the registration technique, and investigate *in-vivo* studies to confirm an acceptable level of accuracy is achieved in an intra-operative setting.

Acknowledgments. The authors wish to thank NSERC and the CIHR Knowledge Translation Fund for supporting this work.

References

1. Grimson, E., Leventon, M.: 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)
2. Fuchs, H., Livingston, M.A., Raskar, R., et al.: Augmented Reality Visualization for Laparoscopic Surgery. In: Wells, W.M., Colchester, A.C.F., Delp, S.L. (eds.) MICCAI 1998. LNCS, vol. 1496, pp. 934–943. Springer, Heidelberg (1998)
3. Su, L.M., et al.: Augmented Reality During Robot-assisted Laparoscopic Partial Nephrectomy: Toward Real-Time 3D-CT to Stereoscopic Video Registration. J. Urology 73(4), 896–900 (2009)
4. Teber, D., et al.: Augmented Reality: A New Tool To Improve Surgical Accuracy during Laparoscopic Partial Nephrectomy? Preliminary In Vitro and In Vivo Results. European Urology 56(2), 332–338 (2009)

5. Cheung, C.L., et al.: Fusion of stereoscopic video and laparoscopic ultrasound for minimally invasive partial nephrectomy. In: Medical Imaging 2009: Visualization, Image-Guided Procedures, and Modeling. Proc. SPIE, vol. 7261, pp. 726109–726110 (2009)
6. Leven, J., et al.: DaVinci canvas: a telerobotic surgical system with integrated, robot-assisted, laparoscopic ultrasound capability. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 811–818. Springer, Heidelberg (2005)
7. Linte, C.A., et al.: Virtual reality-enhanced ultrasound guidance: A novel technique for intracardiac interventions. Comput. Aid. Surg. 13, 82–94 (2008)
8. Lindseth, F., Tangen, G., Lango, T., Bang, J.: Probe calibration for freehand 3D ultrasound. Ultrasound Med. Biol. 29(11), 1607–1623 (2003)
9. Mercier, L., Lango, T., Lindseth, F., Collins, D.L.: A review of calibration techniques for freehand 3-D ultrasound systems. Ultrasound Med. Biol. 31(4), 449–471 (2005)
10. Baghani, A., Eskandari, H., Salcudean, S., Rohling, R.: Measurement of viscoelastic properties of tissue mimicking material using longitudinal wave excitation. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 56(7), 1405–1418 (2009)
11. Poon, T.C., Rohling, R.N.: Tracking a 3-D ultrasound probe with constantly visible fiducials. Ultrasound in Medicine & Biology 33(1), 152–157 (2007)
12. Prager, R.W., Rohling, R.N., Gee, A.H., Berman, L.: Rapid calibration for 3-D freehand ultrasound. Ultrasound Med. Biol. 24, 855–869 (1998)

Automatic Learning Sparse Correspondences for Initialising Groupwise Registration

Pei Zhang, Steve A. Adeshina, and Timothy F. Cootes

Imaging Science and Biomedical Engineering, The University of Manchester, UK
{Pei.Zhang-2,steve.adeshina}@postgrad.manchester.ac.uk,
tim.cootes@manchester.ac.uk

Abstract. We seek to automatically establish dense correspondences across groups of images. Existing non-rigid registration methods usually involve local optimisation and thus require accurate initialisation. It is difficult to obtain such initialisation for images of complex structures, especially those with many self-similar parts. In this paper we show that satisfactory initialisation for such images can be found by a parts+geometry model. We use a population based optimisation strategy to select the best parts from a large pool of candidates. The best matches of the optimal model are used to initialise a groupwise registration algorithm, leading to dense, accurate results. We demonstrate the efficacy of the approach on two challenging datasets, and report on a detailed quantitative evaluation of its performance.

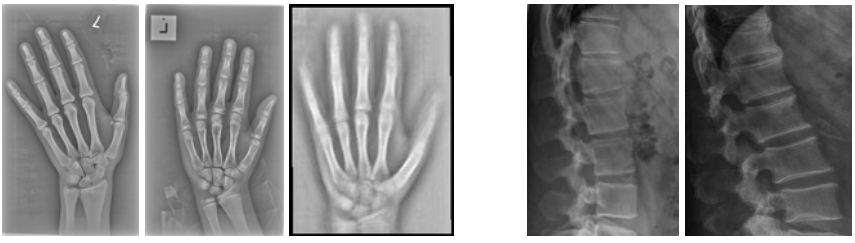
1 Introduction

Groupwise non-rigid image registration is a powerful tool to automatically establish dense correspondences across large sets of images of similar but varying objects. Such correspondences are widely used, for instance to construct statistical models of shape or appearance [1]. These models have a wide range of applications in medical image processing, such as segmentation of anatomical structures or morphometric analysis.

Existing groupwise techniques [2,3,4] generally treat registration as an optimisation problem which is solved with local minimisation methods. As such they are sensitive to initialisation and will fail if "good-enough" starting points are not found. This problem is prominent when registering images of objects with considerable shape variation and multiple similar sub-structures, such as radiographs of the human hand (Fig. 1a). The reason is that non-rigid registration methods usually use an affine transformation to find an approximate initialisation, then refine this with local optimisation. Unfortunately this is insufficient for objects with complex structures because there are too many local minima. For instance, Fig. 1b shows the average of a set of hand radiographs after affine alignment. This shows that the registration failed to register the fingers adequately on many of the examples.

Where objects have a number of distinctive parts, these can be individually located and used to initialise further registration. However, important classes of

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 635–642, 2010.



a) Hand Examples b)Affine mean c) Spine examples

Fig. 1. Examples of hand and spine radiographs used in experiments

object (such as the human hand or spine) contain multiple repeating structures, leading to inherent ambiguity when attempting to match any given part. An effective way to resolve this ambiguity is to use a geometric model of a set of parts, that is, the parts+geometry model [5,6]. Such models are often created so as to optimise their ability to detect or discriminate objects [7]. In this paper we optimise them so as to achieve accurate correspondence.

In [8] we have shown that careful selection of a small number of parts on a single image can be used to construct a parts+geometry model from the whole set, which can lead to good sparse correspondences. These can be used to initialise a dense registration, leading to accurate results. In this paper we describe an algorithm which avoids the manual intervention, and automatically selects a good set of parts for a sparse model, so that the whole process of correspondence establishment is still fully automatic.

The work most similar to ours is that of Langs et al. [9,10]. They describe a method of constructing sparse shape models from unlabelled images, by finding multiple interest points and using minimum description length (MDL) principle to determine optimal correspondences, finding the model which minimises the description of the feature points. Another related approach was developed by Karlsson and Åström [11], who built patch models to minimise an MDL function, estimating the cost of explaining the whole of each image using the patches (by including a cost for the regions not covered by patches).

Both of the above approaches represent shape with a Point Distribution Model [1]. Such representations are useful for local optimisation, but cannot efficiently deal with multiple candidates. By instead learning a parts+geometry model, where the geometry is modelled with a sparse graph, we can take advantage of dynamic programming (DP) algorithms which can efficiently find the global optima where multiple candidates are present. Like Karlsson and Åström, our cost function is based on explaining the whole of an image region, but in our case this is done by constructing a model of the whole image using non-rigid deformation based on the centres of the part models. In addition our goal is somewhat different—we seek a sparse set of parts which can be used to initialise a local optimisation based groupwise registration scheme.

Our main contributions are **1)** automatic selection of a suitable subset of feature parts which are likely to be well localised; **2)** unsupervised learning

of a parts+geometry model which can be used to obtain the optimal sparse correspondences; **3)** a comparison of two different methods of locating the parts.

2 Part Models

We first construct a set of candidate parts which represent structures that are present in most of the images. Given a set of such parts we can automatically construct a geometric model of their relative positions (see Sect. 3). The aim then is to select a subset of parts which leads to the model that determines the best set of dense correspondences between the training images.

We describe the position $\mathbf{x} = (x, y)$, scale s and orientation θ of each part with a pose parameter $\mathbf{p} = \{\mathbf{x}, s, \theta\}$. We have experimented with two methods of representing and localising parts—patch based and SIFT based:

Patch Models: Each part is represented as a statistical model of the intensities over an oriented square region centred at $\mathbf{x}$. Let $\mathbf{g}(I, \mathbf{p})$ be the intensities sampled from the region defined on image I with the pose parameter $\mathbf{p}$, normalised to have a mean of zero and unit variance. The quality of fit to such a model is evaluated as $f_i(\mathbf{g}(I, \mathbf{p})) = \beta \sum_{j=1}^{n} |g_j - \bar{g}_{ij}|/\sigma_{ij}$, where $\bar{\mathbf{g}}_i$ is the vector of mean intensities for the region and σ_{ij} is an estimate of the mean absolute difference from the mean across a training set. β is a normalisation factor chosen so that the standard deviation of the best fits across the training set is unity[1].

Locating candidates for each part involves a multi-resolution search at a range of scales and orientations, in which local optima are located at a coarse scale and refined at finer scales. This approach allows us to quickly search large regions, usually resulting in a few tens of hypotheses.

In order to obtain a set of part models for an unlabelled image set, we arbitrarily choose one image as the reference image[2]. Then we use it to generate a group of patches for a range of sizes, arranged in an overlapping grid pattern (Fig. 2a). We use the region within a given patch to build a part model, and search the rest of the images for the best match on each. We rank the best matches by the quality of fit, and rebuild the model from the best 50% of these. The resulting model is then used to search the images again to find its matches.

To select the models which are likely to have good localisation ability from the set, we sort them by how well the optima is localised [3], and select the best. The set of retained models is denoted as $\mathbf{R}_{\text{sub}}$. Figure 2b shows some examples of the selected part models.

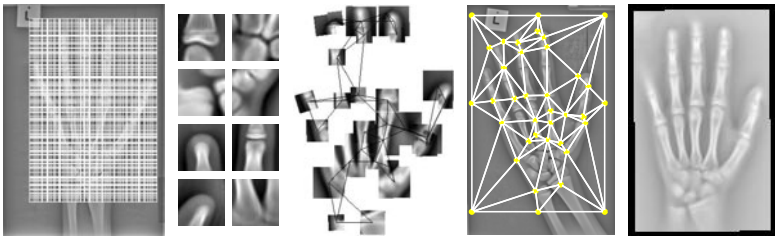
SIFT Models: We can also represent each part using a SIFT signature[4] [12]. To search for candidates for a part we compute interest points using **1)** an edge

[1] We find this form (which assumes the data has an exponential distribution) is more robust than normalised correlation, which is essentially a sum of squares measure.

[2] we find that the algorithm is insensitive to the choice of the reference image.

[3] This is evaluated by $q_i = \min_{\delta\mathbf{x}} \|\bar{\mathbf{g}}(I, \mathbf{p}_i + \delta\mathbf{x}) - \bar{\mathbf{g}}(I, \mathbf{p}_i)\|_2$, where $\bar{\mathbf{g}}(I, \mathbf{p}_i) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{g}(I_k, \mathbf{p}_{i,k})$, $\mathbf{p}_{i,k}$ is the pose of the best match of the part model i on image I_k, $\delta\mathbf{x}$ is the displacement of the model ($1 \leq |\delta\mathbf{x}| \leq 4$ in this paper). If q is small then the model is not good at localising.

[4] http://www.vlfeat.org



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

Fig. 2. (a) An overlapping grid. (b) Typical patch models. (c) A parts+geometry model. (d) Sparse points used for correspondence. (e) The resulting mean reference.

detector and **2)** a variant of a local symmetry point detector [13], which returns local minima of a smoothed edge strength image. We then rank the points using their matches to the part signature and retain the best 50.

To generate a set of initial candidate part models, we locate interest points on every image. We then select those on the reference image which pass a variant of forward-backward matching [14]. A point is selected and added to $\mathbf{R}_{\text{sub}}$ if for at least 50% of the rest of the images in the set, the best match in the other image also matches back to the original point.

3 Parts+Geometry Models

Model Definition: Given a set of m parts from $\mathbf{R}_{\text{sub}}$, we can automatically build a parts+geometry model $\mathcal{G}$ (see below), which can be used to disambiguate the multiple responses of a single part model. An example is given in Fig. 2c. $\mathcal{G}$ represents the object of interest using the parts together with a collection of pairwise geometric relationships between each part, which are defined by a set of arcs ω. For each part, we take its best K matches on each image as its match candidates.

Let $f_i(\mathbf{p})$ be the cost associated with part i having pose $\mathbf{p}$. The relationship between part i and part j can be represented in the cost function, $f_{ij}(\mathbf{p}_i, \mathbf{p}_j)$. This can be derived from the joint pdf of the parameters. In the following we take advantage of the fact that the orientation and scale of the objects are roughly equivalent in each image, and simply use a cost function based on the relative position of the parts, $f_{ij}(\mathbf{p}_i, \mathbf{p}_j) = ((\mathbf{x}_j - \mathbf{x}_i) - \mathbf{d}_{ij})^{\mathrm{T}} \mathbf{S}_{ij}^{-1} ((\mathbf{x}_j - \mathbf{x}_i) - \mathbf{d}_{ij})$, where $\mathbf{d}_{ij}$ is the mean separation of the two parts, and $\mathbf{S}_{ij}$ is an estimate of the covariance matrix. If the objects are likely to undergo significant scaling or orientation changes across the set, more sophisticated function can be used.

To find the best match of $\mathcal{G}$ on an image, we must select one match candidate point $\mathbf{p}_i$ for each part (from its K match candidates) so as to minimise the following function

$$C = \sum_{i=1}^{m} f_i(\mathbf{p}_i) + \alpha \sum_{(i,j)\in\omega} f_{ij}(\mathbf{p}_i, \mathbf{p}_j). \tag{1}$$

The value of α affects the relative importance of part and geometry matches. Given multiple possible candidates for each part, we can use graph algorithms to locate the optimal solutions to (1). We use a method which is similar to that used in [15], in which a network is created where each node can be thought of as having at most two parents. The optimal solution for this can be obtained with a variant of DP algorithm, in $O(mK^3)$ time. If K is modest, this is still fast.

Model Construction: Given a set of parts in the reference image, we can construct a set of connecting arcs, ω, in such a way that each point other than the first two is connected to two parents [8]. The geometric relationships for each arc $(i, j) \in \omega$ are initialised with Gaussians (with standard deviation set to 25% of the length of the arc in the reference image). We then refine the model by applying it to the responses on each image, ranking the results by final fit value (per image), and re-estimating the geometric distributions from the results on the best 50% of the images—essentially a form of robust model building.

Model Evaluation: The match of $\mathcal{G}$ to each image defines a set of sparse correspondences. To evaluate the performance of $\mathcal{G}$ we estimate how effectively a model built from its matches can represent the original image set (a description length). To calculate this we augment the points from each best match and with a set of fixed border points on each image (Fig. 2d). We then align the sets of points and compute the mean. We create a triangulation of the mean and use it to warp each image into the reference frame. We compute the mean intensity in the reference frame (Fig. 2e). Finally, we warp the reference image into the frame of each target image, and compute the sum of absolute differences over the region of interest $U(\mathcal{G}) = \sum_{k=1}^{N-1} \sum_{\mathbf{x} \in R} \left\| I_k(\mathbf{x}) - I_0(W^{-1}(\mathbf{x})) \right\|$, where $I_k(\mathbf{x})$ is the target image intensity at $\mathbf{x}$, R is the region of interest in the target frame and $W(\mathbf{y})$ is the transformation from reference I_0 to target I_k.

Model Selection: Each subset of parts from $\mathbf{R}_{\text{sub}}$ leads to a model, whose quality can be measured as U. Selecting the best subset of parts is thus a combinatoric problem. We solve it using a population based optimisation algorithm, which is similar to the Genetic Algorithm, to find the $\mathcal{G}$ with the minimal U. We create an initial population by randomly sampling subsets from $\mathbf{R}_{\text{sub}}$. For each set we generate a $\mathcal{G}$ then evaluate it. We then rank the members of the population (each a candidate set of parts) by U. We discard the worst 50%, and generate new candidates from pairs of candidates randomly selected from the best 50%. To generate a new subset we simply randomly sample from the union of parts from the two candidate parent sets. Repeating the above process leads to the best $\mathcal{G}$, whose best matches define the optimal set of sparse correspondences.

4 Establishing Dense Correspondences

To obtain an accurate dense registration, we use the sparse correspondences to initialise a non-rigid registration algorithm [4]. We use a coarse to fine algorithm to improve efficiency and robustness. The approach is to compute a robust mean (Fig. 3b) using the current estimate of the deformation field, then to refine the correspondences to improve the match to this mean (Fig. 3c).

5 Experiments

To demonstrate the approach, we applied it to two different datasets: **1)** 94 radiographs of the hands of children (aged between 11-13), taken as part of study into bone ageing (Fig. 1a). Each image has been marked with 37 points by a human expert; **2)** 106 radiographs of the lumbar spine (Fig. 1b). Each image has 337 manual landmarks placed around the outline of the vertebrae. Both sets of images have a resolution of 0.2mm per pixel and a height of about 1300 pixels.

By systematically initialising patch models in an overlapping grid on the reference image at a range of sizes, we automatically constructed over 1900 and 500 part models for the hand and spine sets respectively. We ranked by their localisability and selected the best 500 for the hands and the best 100 for the spines. We used the population based optimisation to select the best parts+geometry models from the two sets, for a range of different numbers of parts. An example of the resulting parts+geometry models for the hand set is shown in Fig. 3a. The resulting correspondences were used to estimate an initial dense correspondence, from which a robust mean was estimated (Fig. 3b). Groupwise non-rigid registration was applied, giving the final result shown in Fig. 3c. Equivalent results for the spine set are shown in Fig. 3d-f.

We also repeated the experiment on the hands using the SIFT based models.

To evaluate the accuracy of the result we compare with a manual annotation. We used the resulting dense correspondences to warp each set of manual landmarks into a reference frame, computing their mean. We then projected this mean back to each individual image (Fig. 4) and calculated the mean absolute difference between each mean point and the original manual annotation. For the spines, we calculated the mean distance between the mean points and the curve formed by the original manual landmarks.

Table 1 shows the statistics of the resulting point location errors for different numbers of parts for the two datasets. There is not a clear relationship between the number of parts and performance—once sufficient parts are available to cover the main components of the object, adding further elements is unlikely to improve performance and may lead to a decline. We found that the patch based

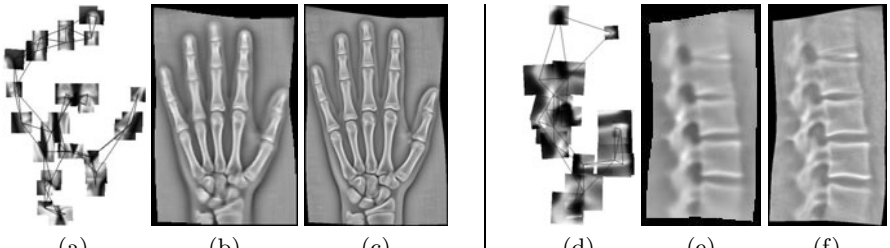


Fig. 3. Examples of the best parts+geometry models and registered means

Table 1. Point location errors (mm) of the dense correspondence

a) Hand set

Parts	SIFT based			Patch based		
	Mean	Med.	90%	Mean	Med.	90%
5	4.2	3.6	6.5	3.3	2.4	7.8
10	2.0	1.2	3.7	2.2	1.7	4.0
15	1.7	1.1	2.9	1.1	1.0	1.6
20	1.4	1.1	2.3	1.3	0.9	2.5
25	1.1	0.9	1.9	1.1	0.9	1.7
30	1.6	1.3	2.9	**1.0**	**0.8**	**1.4**

b) Spine set (patch based only)

Parts	Mean	Med.	90%
5	6.8	3.1	21.9
10	5.2	2.5	13.2
15	5.0	3.1	9.7
20	6.2	3.9	14.5

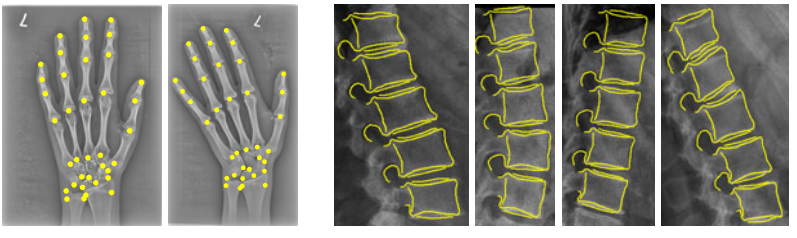


Fig. 4. Examples of projection of average points onto individual image

part models generally lead to better results than their SIFT based counterparts, though locating them is more computationally expensive.

Computing the equivalent error for the hands after performing a standard affine initialisation gives a median error of 12.0mm, which is little improved by further dense registration due to local minima. The parts+geometry method gives substantially better results, demonstrating the importance of good initialisation. Comparison with other published approaches is not easy. The most similar approach in the literature [10] evaluates on a set of only 20 hand radigraphs (0.34mm/pixel), obtaining a mean error of 2.0mm and a median of 1.7mm (though on a larger set of points)—our method appears to give significantly better results.

6 Conclusions and Future Work

We have described an approach for automatically locating dense correspondences across a set of images, by using the sparse matches of a parts+geometry model to initialise groupwise non-rigid registration. It is able to achieve good results on two challenging datasets. The technique potentially can be used for a wide range of other datasets.

In the above we show results for particular choices of numbers of parts. Potentially it will be possible to use the framework to estimate the optimal number of parts (as is done by [11]). However, our objective is not to explain the whole image with the parts+geometry model, just to obtain good enough correspondence

for the dense registration. The most efficient method of selecting the appropriate number is currently under investigation. Moreover, we have used a relatively simple topology for the geometric model in order to allow efficient global optimisation. In future we will try using more highly connected graphs.

References

1. 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)
2. Baker, S., Matthews, I., Schneider, J.: Automatic construction of active appearance models as an image coding problem. IEEE TPAMI 26(10), 1380–1384 (2004)
3. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)
4. Cootes, T.F., Twining, C.J., Petrović, V., Schestowitz, R., Taylor, C.J.: Groupwise construction of appearance models using piece-wise affine deformations. In: Proc. of British Machine Vision Conference, vol. 2, pp. 879–888 (2005)
5. Felzenszwalb, P.F., Huttenlocher, D.P.: Pictorial structures for object recognition. International Journal of Computer Vision 61(1), 55–79 (2005)
6. Donner, R., Micusik, B., Langs, G., Szumilas, L., Peloschek, P., Friedrich, K., Bischof, H.: Object localization based on markov random fields and symmetry interest points. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part II. LNCS, vol. 4792, pp. 460–468. Springer, Heidelberg (2007)
7. Fergus, R., Perona, P., Zisserman, A.: Weakly supervised scale-invariant learning of models for visual recognition. International Journal of Computer Vision 71, 273–303 (2007)
8. Adeshina, S.A., Cootes, T.F.: Constructing part-based models for groupwise registration. In: Proc. of International Symposium on Biomedical Imaging (2010)
9. Langs, G., Peloschek, P., Donner, R., Bischof, H.: Annotation propagation by MDL based correspondences. In: Proc. of Computer Vision Winter Workshop (2006)
10. Langs, G., Donner, R., Peloschek, P., Bischof, H.: Robust autonomous model learning from 2D and 3D data sets. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 968–976. Springer, Heidelberg (2007)
11. Karlsson, J., Åström, K.: MDL patch correspondence on unlabeled images with occlusions. In: Proc. of Computer Vision and Pattern Recogition (2008)
12. Lowe, D.: Object recognition from scale invariant features. In: Proc. of International Conference on Computer Vision, vol. 2, pp. 1150–1157 (1999)
13. Donner, R., Micusik, B., Langs, G., Bischof, H.: Sparse MRF appearance models for fast anatomical structure localisation. In: Proc. of British Machine Vision Conference, vol. 2, pp. 1080–1089 (2007)
14. Fua, P.: A parallel stereo algorithm that produces dense depth maps and preserves image features. Machine Vision and Applications 6, 35–49 (1993)
15. Felzenszwalb, P.F., Huttenlocher, D.P.: Representation and detection of deformable shapes. IEEE TPAMI 27(2), 208–220 (2005)

Hierarchical Multimodal Image Registration Based on Adaptive Local Mutual Information

Dante De Nigris[1], Laurence Mercier[2], Rolando Del Maestro[3], D. Louis Collins[2], and Tal Arbel[1]

[1] McGill University, Centre for Intelligent Machines
{dante,tal}@cim.mcgill.ca
[2] McGill University, Dept. of Biomedical Engineering
{laurence,louis}@bic.mni.mcgill.ca
[3] Montreal Neurological Institute and Hospital, McGill University
rolando.del_maestro@mcgill.ca

Abstract. We propose a new, adaptive local measure based on gradient orientation similarity for the purposes of multimodal image registration. We embed this metric into a hierarchical registration framework, where we show that registration robustness and accuracy can be improved by adapting both the similarity metric and the pixel selection strategy to the Gaussian blurring scale and to the modalities being registered. A computationally efficient estimation of gradient orientations is proposed based on patch-wise rigidity. We have applied our method to both rigid and non-rigid multimodal registration tasks with different modalities. Our approach outperforms mutual information (MI) and previously proposed local approximations of MI for multimodal (e.g. CT/MRI) brain image registration tasks. Furthermore, it shows significant improvements in terms of mTRE over standard methods in the highly challenging clinical context of registering pre-operative brain MRI to intra-operative US images.

Keywords: multimodal image registration, image-guided neurosurgery.

1 Introduction

Traditional multimodal image registration techniques rely on the assumption that features exposed in one modality (e.g. blobs, borders) will be similarly observable in a second modality, although possibly with a different intensity response. In this paper, we consider contexts where this assumption is fundamentally violated, as in image-guided neurosurgery, where the objective is to match pre-operative Magnetic Resonance Images (MRI) of patient brains to intra-operative ultrasound (US) images. Inherently, the two modalities convey very different information. US images indicate changes in acoustical impedance in the direction of the sound wave and therefore expose tissue boundaries and liquid-filled structures relatively well. They do not, however, provide distinctive intensity maps relating to different tissue types, as in MRI. This problem, immediately observable in Figure 1, poses serious challenges to most intensity based techniques, including traditional and

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 643–651, 2010.

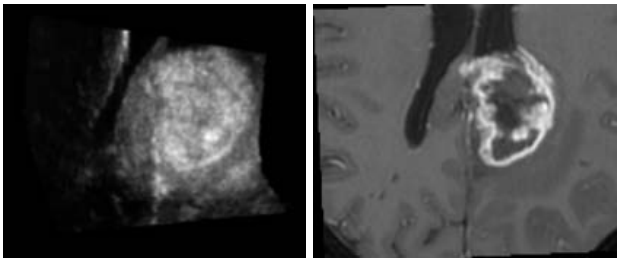


Fig. 1. US brain image acquired during neurosurgery and corresponding pre-operative MRI slice. Notice that the tumor appears as a noisy, diffuse bright blob in the US. In the MRI, the tumor has a bright boundary and a dark center.

normalized mutual information (MI) [1,2,3]. Previous work on MRI/US registration with intensity-based similarity metrics has focused on pre-processing the images to improve their intensity correspondences [4,5] and/or building a mask of the US image where an intensity-based similarity metric has a higher probability of finding correspondence[6].

In this work, we propose a registration framework which embeds a new localized similarity metric expressed as an orientation similarity measure based on a local approximation to MI. Other local MI metrics have been devised [7,8], however, our metric is adaptive, in that it can loosen or tighten its constraints depending on the modalities being registered and the scale at which the images are being registered. We also propose a computationally efficient method for estimating gradient orientations. Experiments with real multimodal (e.g. Computed Tomography (CT) and MRI) brain image registration tasks indicate that it outperforms both standard MI and other established orientation similarity metrics. Furthermore, it shows significant improvements over standard Normalized MI (NMI) in the highly challenging clinical context of registering pre-operative brain MRI to intra-operative US.

2 Method

2.1 Local Mutual Information - Previous Work

In [7], the authors developed a local similarity metric based on the analytical limit of MI as the window of observation approaches the voxel size. The expression for local MI (LMI) is obtained by modeling both images with a first-order Taylor expansion and is a monotonically decreasing function of the angle, θ, between the gradient orientations,

$$LMI(\theta) = C_d + \log_2 |\sin(\theta)| \tag{1}$$

where C_d is a constant that depends on the dimension of the image.

It is not possible to build an energy function by simply summing Eqn. 1 over a set of points in the image, since an extremum would appear whenever any of the points has a minimal inner angle. To circumvent this issue, the authors simplify the expression to,

$$LMI_2(\theta) = \frac{1}{2}\cos^2(\theta) \tag{2}$$

which has a smoother shape and whose energy function exhibits an extremum only when there is a collective coherence in terms of orientation similarity.

In related work [8], the authors preserve the dynamics of the original expression and eliminate the singularity by including an ϵ factor,

$$LMI_3(\theta) = \log_2(\epsilon + |\sin(\theta)|) \tag{3}$$

The two localized metrics have significantly different *coherence-selectivity* trade-offs. A coherent metric, such as Eqn. 2, can be loosely defined as one that is maximal when a majority of points show some degree of image correspondence. In contrast, a selective metric, such as Eqn. 3, has stricter correspondence constraints and therefore exhibits an extremum as soon as a few points comply with such constraints. A measure that favors coherence will tend to be smoother but less accurate because it will effectively average over all sampled pixels.

2.2 Adaptive Local Mutual Information

Our work builds on previous work [7,8] in which MI is locally approximated by a metric appropriately derived from the energy function. We propose an adaptive local orientation-based similar metric (ALMI) that addresses the coherence-selectivity trade-off directly as follows:

$$ALMI(\theta; K, \theta_c) = 2 - \frac{1}{1 + e^{-K(\theta - \theta_c)}} - \frac{1}{1 + e^{-K(\pi - \theta - \theta_c)}} \tag{4}$$

The selectivity of this smooth sigmoid-based function is adaptable by varying the curvature, K, and cutoff angle, θ_c. Figure 2 illustrates how the two previously proposed measures compare to various configurations of our proposed expression. An important advantage of ALMI is that it saturates smoothly to a maximum as $\theta \to \{0, \pi\}$. This property allows the energy function to reach a stable state as the optimizer approaches an extremum.

2.3 Adaptive Selectivity and Sampling

In addition to the metric proposed above, we propose an adaptive multiscale pixel selection scheme and a strategy to parametrize the similarity metric. Rather than adopting a standard low-pass image pyramid [9] in which the images are both blurred and down-sampled at each level, we use a smoothing image pyramid in which only the scale of Gaussian blurring is increased at each level. Figure 3 illustrates how the gradient orientations change with different blurring scales.

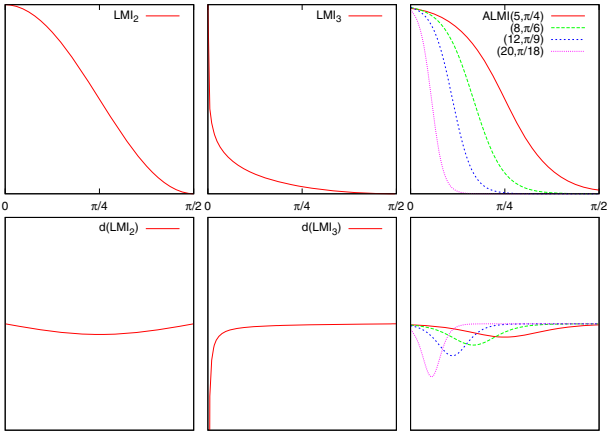


Fig. 2. The top row shows the two previous proposed localized measures and our proposed measure as a function of θ. The bottom row shows their respective derivatives.

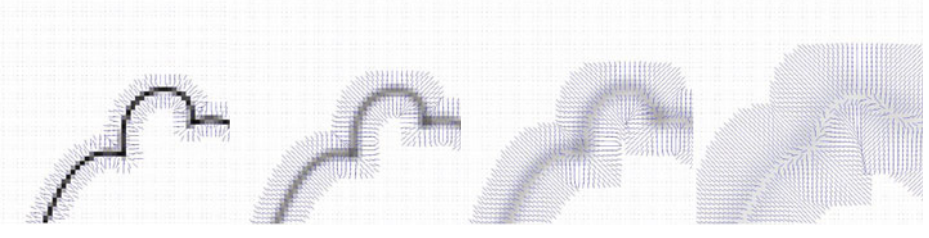


Fig. 3. Gradient Orientations (short blue lines) at different blurring scales, $\sigma \in 0, 1, 2, 4$

As the scale of diffusion is increased, the spatial support of a structure (i.e. the extent of image area where the gradient orientations accurately indicate the orientation of a structure of reference) tends to increase proportionately to the ratio of the current scale to the previous scale.

Further performance gains can be attained by focusing computations on a cleverly selected subset of voxels in the images. In this work, our conjecture is that the gradient magnitude is a valid indicator of the *reliability* of an estimated orientation and therefore we select high gradient voxels. The performance of our method is improved by adapting this sampling mask, as well as the parameters of the ALMI similarity metric in relation to the current Gaussian scale and the quality of the modalities being registered. We express the adaptive masking function $M(\mathbf{x})$ as,

$$M(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad |\nabla I_f(\mathbf{x})| > g(p_\sigma) \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

where $p_\sigma = \left(\frac{\sigma}{\sigma_1}\right) p_{\sigma_1}$ and $g(p)$ is the threshold value that captures the top p percentile of the fixed image gradient magnitude.[1]

2.4 Cost Saving Approximations

A metric based on gradient orientations incurs a computational cost related to evaluating the gradient at each point of interest. We propose a simplifying implementation in which the gradients are computed only once for both the fixed and the moving image. Subsequent gradient orientations of the transformed moving image are estimated from the initial orientations.

In order to permit non-rigid registrations, we make the simplifying assumption that the transformation in the voxel neighborhood can be well approximated by a locally rigid transformation. Hence, we can estimate the gradient orientation of a point by multiplying the gradient orientation of the point prior to transformation with an estimated rotation matrix as follows:

$$\begin{aligned} \theta(\mathbf{x}) &= \angle(\nabla I_f(\mathbf{x}), \nabla I_m(T(\mathbf{x}))) \\ &\approx \angle(\nabla I_f(\mathbf{x}), R \cdot \nabla I_m(\mathbf{x}' = T(\mathbf{x}))) \end{aligned}$$

Such a scheme eliminates the effect of intensity-interpolation artifacts and minimizes the expense of using high quality gradient operators.

3 Experiments

In this section, we present the results of evaluating our method, implemented within the Elastix toolbox [10], with two clinical datasets. Section 3.1 details the results of CT/MR rigid registration cases from a public dataset, while Section 3.2 details the results of rigid and non-rigid registrations with real MR and US images during image-guided neurosurgery.

3.1 Rigid Registration of CT/MR Brain Images

We evaluated the performance of our method with a series of CT/MR brain registration cases from the Retrospective Image Registration Evaluation (RIRE)[2] project. Since both modalities preserve structure and a large extent of that structure has low curvature, we chose a selective profile ($K = 20$, $\theta_c = \pi/18$) for all resolutions.

Table 3.1 lists the results of our proposed method with an adaptive masking techniques ($ALMI_m$) and without ($ALMI$). We also show the results of registration with previous proposed local metrics based on orientation similarity (LMI_2,LMI_3), a standard MI technique, a standard NMI technique and a

[1] In our experiments, we found that setting $p_{\sigma_1} = 0.1$ (i.e. sampling from the top 10% at the finest resolution) consistently yielded very good results.

[2] http://www.insight-journal.org/rire

more sophisticated technique which maximizes NMI with Generalized Partial Volume Interpolation (NMIGPV) [11].[3] The mTRE is evaluated over the Volumes of Interest (VOI) of all cases. ALMI-based registrations on average show an improved accuracy over the previous proposed orientation similarity metrics. In addition, the metrics with high selectivity generally perform better than the coherent metric, LMI_2. The standard MI and NMI methods exhibit relatively large mTREs, possibly because they do not embed any scheme to reduce the effect of interpolation-induced artifacts [12,3]

Table 1. Registration Results expressed in mTRE (in mm) of the VOIs of six patients

	MI	NMI	NMIGPV	LMI_2	LMI_3	$ALMI$	$ALMI_m$
	Intensity Based			Orientation Based			
CT/T1	3.20	3.43	0.87	2.31	1.42	1.14	1.00
CT/T2	1.90	5.03	1.11	2.87	1.56	1.45	1.39
CT/PD	4.11	3.88	0.90	2.36	0.82	1.07	0.91

3.2 MRI/US for Image Guided Neurosurgery

We apply our framework to the clinical context of image-guided neurosurgery where intra-operative US images are acquired during open craniotomies and matched to pre-operative MRI in order to correct for brain shift (linear and non-linear). Figure 1 shows extracted planes from a registered US and MR volume. Notice that apart from key structural components like the right lateral ventricle and the fissure, the ultrasound does not have a global intensity or orientation mapping with respect to the MR slice. However, we can argue that the US high gradient pixels do exhibit a loose orientation correspondence with the corresponding MR pixels.

Figure 4 illustrates our proposed similarity metric evaluated with four different configurations formed by a choice of a selective ALMI or a non-selective ALMI, and a choice of full sampling or sampling of the high gradient regions. The energy function between the US and MR slice of Figure 1 is evaluated in the parametric space of translational displacements in the x and y axis. We can observe that choosing a non-selective ALMI greatly reduces the number of local extrema. Furthermore, sampling only the high gradient regions of the US further reduces the number of local extrema, increasing the probabilities of a successful registration. Hence, we implement a strategy based on ALMI that relaxes the orientation similarity constraint as finer resolutions are approached.

Multimodal image registration experiments are performed on four tumor cases acquired at the Montreal Neurological Institute. The cases include: 1) high-grade glioma in the left frontal lobe, 2) high-grade glioma in the left occipital lobe,

[3] The results of the rigid CT/MR registration were obtained directly from the RIRE website, where fellow researchers upload the results from their implementations using the same real datasets. We chose to compare our method against those results so as to avoid biases introduced by implementing other techniques.

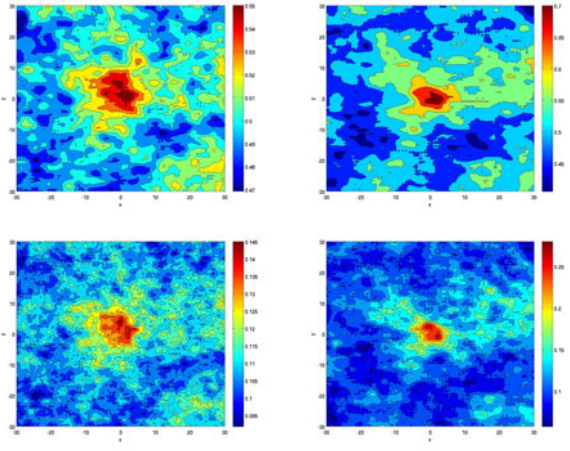


Fig. 4. The energy functions in the top row use a relaxed orientation matching constraint, evaluated with ($K = 5$, $\theta_c = \pi/4$) while the energy functions on the bottom row use a strict orientation matching constraint, evaluated with ($K = 5$, $\theta_c = \pi/4$). The energy functions in the left column evaluate all overlapping pixels, while the energy functions on the right column evaluate only the top 10% high gradient US pixels.

3) high-grade glioma in the right parietal lobe, and 4) low-grade glioma in the left temporal lobe. Each case includes a pre-operative gadolinium- enhanced T1-weighted MRI and a 3D ultrasound image acquired on the dura before starting the resection. Ultrasound images were acquired with an HDI 5000 (ATL/Philips, USA) using a P7-4 MHz phased array transducer at a depth setting of 6.5 or 8 cm. Each acquisition consisted in 200-600 tracked images that were reconstructed into a 3D volume at a resolution of 0.3 x 0.3 x 0.3 mm. To validate registration, an expert (neurosurgeon) chose 7 to 11 corresponding anatomical points in both modalities. With every estimated transformation, we transform the locations of the anatomical points in the modality of the fixed image (US), evaluate the distance (i.e. error) from the corresponding anatomical points in the moving image (MR), and compute their mean.

Table 2 shows the results of the four cases after a rigid registration and after a subsequent non-rigid registration characterized by a B-Spline transformation. We compared our strategy against a standard multiresolution registration that maximizes NMI, and against a state-of-the-art method that pre-processes the US and MR volumes and then maximizes NMI [6]. Both NMI based registration techniques have relatively high registration errors in this domain. Only in cases 3 and 4, where there was only a small displacement, did the rigid registration show improvement. Non-rigid registrations led to consistent divergence from the ground truth. Conversely, our proposed approach manages to reduce the mTRE in all cases, three of them with an average accuracy below 2 mm. Furthermore, applying a non-rigid transformation further improved the accuracy in the first two cases, and slightly degraded the second two cases.

Table 2. MRI/US Registration Results expressed in mTRE (in mm) of expert selected anatomical points of interest

Case	Initial	NMI		Pre-processing + NMI		ALMI	
		Rigid	Nonrigid	Rigid	Nonrigid	Rigid	Nonrigid
1	10.50	14.75	17.12	16.03	13.81	1.87	1.40
2	10.86	17.52	20.80	13.94	13.94	2.96	2.53
3	2.62	1.98	5.70	2.31	4.28	1.63	1.86
4	3.24	2.91	7.44	3.66	8.07	1.93	2.06

4 Discussion

We have presented a new approach for multimodal image registration based on an adaptive localized similarity metric and an adaptive image masking scheme, We have provided basic guidelines for the choice of ALMI parameters in relation to the quality of the modalities being registered and the Gaussian blurring scale. Nevertheless, in an extended version of this work, we will seek to provide a more detailed analysis of the influence of different parameter selections. We also introduced a computationally efficient implementation that evaluates the image gradients only at the start of each resolutions optimization. We have shown improvements in accuracy and robustness across various modalities. We have also demonstrated that the approach is general enough to tackle a difficult problem such as MRI/US registration.

References

1. Maes, F., Collignon, A., Dirk, V., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imaging 16, 187–198 (1997)
2. Viola, P., Wells, W.M.: Alignment by maximization of mutual information. In: IJCV, pp. 16–23 (1995)
3. Pluim, J., Maintz, J., Viergever, M.: Mutual-information-based registration of medical images: a survey. IEEE Trans. Med. Imaging 22(8), 986–1004 (2003)
4. Arbel, T., Morandi, X., Comeau, R., Collins, D.L.: Automatic non-linear MRI-Ultrasound registration for the correction of intra-operative brain deformations. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, pp. 913–922. Springer, Heidelberg (2001)
5. 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. Med. Imaging 20, 1038–1049 (2001)
6. Ji, S., Wu, Z., Hartov, A., Roberts, D.W., Paulsen, K.D.: Mutual-information-based image to patient re-registration using intraoperative ultrasound in image-guided neurosurgery. Medical Physics 35(10), 4612–4624 (2008)
7. Karaçali, B.: Information theoretic deformable registration using local image information. IJCV 72, 219–237 (2007)

8. Biesdorf, A., Wörz, S., Kaiser, H.J., Rohr, K.: Hybrid spline-based multimodal registration using local measures for joint entropy and mutual information. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 607–615. Springer, Heidelberg (2009)
9. Lindeberg, T.: Scale-space theory in computer vision. Springer, Heidelberg (1993)
10. Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J.: elastix: a toolbox for intensity-based medical image registration. IEEE Trans. Med. Imaging 29(1), 196–205 (2010)
11. Chen, H., Varshney, P.K.: Mutual information-based CT-MR brain image registration using generalized partial volume joint histogram estimation. IEEE Trans. Medical Imaging 22(9), 1111–1119 (2003)
12. Tsao, J.: Interpolation artifacts in multimodality image registration based on maximization of mutual information. IEEETrans. Med. Imaging 22(7), 854–864 (2003)

LogDemons Revisited: Consistent Regularisation and Incompressibility Constraint for Soft Tissue Tracking in Medical Images

T. Mansi, X. Pennec, M. Sermesant, H. Delingette, and N. Ayache

INRIA-Méditerranée, Asclepios Project, Sophia Antipolis, France

Abstract. Non-linear image registration is a standard approach to track soft tissues in medical images. By estimating spatial transformations between images, visible structures can be followed over time. For clinical applications the model of transformation must be consistent with the properties of the biological tissue, such as incompressibility. LogDemons is a fast non-linear registration algorithm that provides diffusion-like diffeomorphic transformations parameterised by stationary velocity fields. Yet, its use for tissue tracking has been limited because of the *ad-hoc* Gaussian regularisation that prevents implementing other transformation models. In this paper, we propose a mathematical formulation of demons regularisation that fits into LogDemons framework. This formulation enables to ensure volume-preserving deformations by minimising the energy functional directly under the linear divergence-free constraint, yielding little computational overhead. Tests on synthetic incompressible fields showed that our approach outperforms the original logDemons in terms of incompressible deformation recovery. The algorithm showed promising results on one patient for the automatic recovery of myocardium strain from cardiac anatomical and 3D tagged MRI.

1 Introduction

Tissue tracking in sequences of medical images is an important task in many applications, either for therapy guidance or diagnosis. However there is no easy way to achieve it, even interactively. A standard approach is now to use non-linear image registration to estimate the spatial transformation between different images, for instance brain shift [7]. However, for such clinical applications one-to-one mapping must often be ensured and the model of transformations that is used must be consistent with the properties of the tissue to track. In particular, constraining the registration to be volume-preserving showed great improvement when tracking incompressible tissues such as the heart [3,9].

Mathematical frameworks based on diffeomorphic deformations [2,4] have been developed to get one-to-one mappings between the images to register. Among them, logDemons [14] is an efficient demons-based registration algorithm that provides diffeomorphic transformations parameterised by stationary velocity fields. However, although mathematical justifications of demons optimisation have been provided [14], theoretical foundations of the Gaussian regularisation

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 652–659, 2010.

still has to be consolidated [11], which makes the algorithm difficult to adapt to other deformation models.

Several approaches for incompressible image registration have been proposed. A first method is to constrain the Jacobian determinant of the transformations to equal 1. However, this constraint is computationally demanding due to its non-linearity [12]. Linear approximations have been proposed [3] but volume drifts may appear at large deformations. Velocity fields can be made incompressible by constraining them to be divergence-free. Thus, incompressible fluid registration is achieved by projecting the update velocity onto the space of divergence-free vector fields [13]. Nevertheless, the fluid model might not be appropriate for tracking biological soft tissues like myocardium. In [9], we proposed an incompressible demons algorithm where the update velocity field was made divergence-free using Helmholtz decomposition. However, the approach was suboptimal as the constraint did not consider demons Gauss-Newton minimisation space. Furthermore, volume drifts were controlled using the linear approximation of the condition on deformations, which does not hold on large deformations.

This paper presents an efficient and consistent framework for demons-based incompressible registration. We first propose a mathematical justification of the Gaussian regularisation, which enables to integrate the incompressibility constraint seamlessly by working on the space of divergence-free velocity fields. The main advantages of this are: i) the constraint is linear with little computational overhead, ii) the parameter of the deformations are constrained: no volume drifts appear, iii) the transformation minimises a constrained energy functional: the optimal incompressibility field according to the logDemons minimisation scheme is found. The algorithm was validated against synthetic data and applied on clinical cardiac MRI to estimate 3D myocardium motion.

2 Methods

2.1 Background: Log-Domain Diffeomorphic Demons

LogDemons algorithm estimates a dense non-linear transformation ϕ that best aligns a template image T to a reference image R [14]. ϕ belongs to the space generated by the one-parameter subgroups of diffeomorphisms $\mathbb{G}$. They are parameterised by stationary velocity fields $\mathbf{v}$ through the exponential map $\phi = \exp(\mathbf{v})$ [1]. The images R and T are registered by minimising in the space of velocities, called *log-domain*, the energy functional: $\mathcal{E}(\mathbf{v}, \mathbf{v}_c) = 1/\sigma_i^2 \ \|R - T \circ \exp(\mathbf{v}_c)\|_{L_2}^2 + 1/\sigma_x^2 \ \|\log(\exp(-\mathbf{v}) \circ \exp(\mathbf{v}_c))\|_{L_2}^2 + 1/\sigma_d^2 \ \|\nabla \mathbf{v}\|^2$, where σ_i^2 relates to the noise in the images and σ_d^2 controls the regularisation strength. This equation is function of two variables: The velocity field $\mathbf{v}$ parameterises the transformation ϕ to recover whereas $\mathbf{v}_c$ parameterises an intermediate transformation $\phi_c = \exp(\mathbf{v}_c)$ that models the *correspondences* between the voxels of the two images. During the *optimisation step*, $\mathcal{E}(\mathbf{v}, \mathbf{v}_c)$ is minimised with respect to $\mathbf{v}_c$. This amounts to finding the optimal matching between R and T without considering the regularisation. Under the diffeomorphic update rule $\phi_c \leftarrow \phi \circ \exp(\delta \mathbf{v})$, the optimal update velocity writes $\delta \mathbf{v}(\mathbf{x}) = -(R(\mathbf{x}) - T \circ \phi(\mathbf{x}))/(\|J(\mathbf{x})\|^2 + \sigma_i^2/\sigma_x^2) J(\mathbf{x})$.

$J(\mathbf{x})$ is the symmetric gradient $J(\mathbf{x}) = (\nabla R(\mathbf{x}) + \nabla(T \circ \phi)(\mathbf{x}))/2$. The correspondence velocity $\mathbf{v}_c$ is then updated using the first order approximation of Baker-Campbell-Hausdorff (BCH) formula $\mathbf{v}_c = Z(\mathbf{v}, \delta\mathbf{v}) = \mathbf{v} + \delta\mathbf{v} + 1/2[\mathbf{v}, \delta\mathbf{v}] + 1/12[\mathbf{v}, [\mathbf{v}, \delta\mathbf{v}]] + O(\|\delta\mathbf{v}\|^2)$, where the Lie bracket $[\cdot, \cdot]$ is defined by $[\mathbf{v}, \delta\mathbf{v}] = (\nabla\mathbf{v})\delta\mathbf{v} - (\nabla\delta\mathbf{v})\mathbf{v}$. This approach has experimentally shown promising results in terms of image registration and statistics on diffeomorphisms [4,14]. Finally, the *regularisation step* estimates the optimal regularised transformation ϕ by minimising $\mathcal{E}(\mathbf{v}, \mathbf{v}_c)$ with respect to $\mathbf{v}$, which is approximated by smoothing the correspondence velocity $\mathbf{v}_c$ with a Gaussian kernel G_σ. However, how G_σ relates to $\mathcal{E}(\mathbf{v}, \mathbf{v}_c)$ remains to be consolidated [11].

2.2 Revisiting Demons Gaussian Regularisation

A consistent mathematical formulation of demons regularisation is required to integrate incompressibility in logDemons. In [10], the authors demonstrate that Gaussian filtering solves the Tikhonov estimation problem with equal weighting of the spatial derivatives in the Taylor series sense. We thus replace the logDemons regulariser $\|\nabla\mathbf{v}\|^2$ by the Tikhonov regulariser to get:

$$\mathcal{E}_{reg}(\mathbf{v}) = \frac{1}{\sigma_x^2}\|\log(\exp(-\mathbf{v}) \circ \exp(\mathbf{v}_c))\|_{L_2}^2 + \int_\Omega \sum_{k=1}^{+\infty} \left(\sum_{i_1+\ldots+i_k=k} \frac{\|\partial_{i_1..i_k}\mathbf{v}\|^2}{\sigma_x^2 \sigma_d^{2k} k!} \right)$$

In this equation, Ω is the image domain and $\partial_{i_k..i_l}$ denotes the composition of spatial derivatives $\partial_{i_k}..\partial_{i_l}$. The parameter σ_x^2 has been introduced into the regulariser to simplify calculations. More importantly, the regularisation weight σ_d^2 is now function of the derivative orders to preserve the shape of the impulse response related to the regulariser [10]. The previous equation is minimised by linearising its first term using the zeroth order approximation of BCH formula, $\log(\phi^{-1} \circ \phi_c) = \mathbf{v}_c - \mathbf{v}$. The resulting equation, which is exactly a Tikhonov estimation problem, is minimised in the Fourier domain. The optimal velocity field $\mathbf{v}$ verifies $\sum_{k=0}^{\infty}(\mathbf{w}^T\mathbf{w})^k/(\sigma_d^{2k}k!)\hat{\mathbf{v}}(\mathbf{w}) = \exp(\mathbf{w}^T\mathbf{w}/\sigma_d^2)\,\hat{\mathbf{v}}(\mathbf{w}) = \hat{\mathbf{v}}_c(\mathbf{w})$, which is exactly demons Gaussian smoothing G_σ, with $\sigma^2 = 2/\sigma_d^2$. The width of the Gaussian kernel corresponds to the strength of the regularisation.

This formulation is built up on two key elements. First, the correspondence field ϕ_c decouples the regularisation from the optimisation, making the minimisation of the registration energy independent of the optimisation. Second, the coupling term $\|\log(\phi^{-1} \circ \phi_c)\|^2$ is approximated as a least-square problem to get the Gaussian smoothing. These two elements must be ensured in any demons algorithm to justify the Gaussian regularisation.

2.3 Incompressible LogDemons

The proposed regulariser enables to integrate the incompressibility constraint into the algorithm. A transformation ϕ is locally incompressible if its Jacobian determinant $\det(\nabla\phi) = 1$. In fluid dynamics, one uses the infinitesimal version: a fluid is said incompressible if the divergence of its velocity $\mathbf{v}$ is null. For

diffeomorphic transformations one can show that the converse is true: integrating divergence-free velocities over time yields incompressible deformations [6]. Making LogDemons incompressible thus consists in constraining the velocity fields $\mathbf{v}$ to be divergence-free. Demons optimisation step is not modified, as it optimises $\mathbf{v}_c$ only, but demons regularisation energy is now optimised under the divergence-free constraint, which amounts to minimising the Lagrange function:

$$\mathcal{P}(\mathbf{v},p) = \frac{1}{\sigma_x^2}\|\mathbf{v}_c - \mathbf{v}\|_{L_2}^2 + \sum_{k=1}^{+\infty}\left(\sum_{i_1+\ldots+i_k=k}\frac{\|\partial_{i_1..i_k}\mathbf{v}\|^2}{\sigma_x^2\sigma_d^{2k}k!}\right) - \frac{2}{\sigma_x^2}\int_\Omega p\,\nabla\cdot\mathbf{v} \quad (1)$$

In this equation, the Lagrange multiplier p is a scalar function of the Sobolev space $H_0^1(\Omega)$ that vanishes at infinity. Optima of (1) are found by solving $\partial_{\mathbf{v}}\mathcal{P}(\mathbf{v},p) = 0$:

$$\mathbf{v} + \sum_{k=1}^{\infty}\frac{(-1)^k}{\sigma_d^{2k}k!}\Delta^k\mathbf{v} = \mathbf{v}_c - \nabla p \quad (2)$$

with $p = 0$ at the domain boundaries $\partial\Omega$. The divergence of (2) under the optimal condition $\nabla\cdot\mathbf{v} = 0$ yields $\Delta p = \nabla\cdot\mathbf{v}_c$ with 0-Dirichlet boundary conditions, which can be solved independently of $\mathbf{v}$ to get p. $\mathbf{g} = \mathbf{v}_c - \nabla p$ is thus the L_2 projection of $\mathbf{v}_c$ onto the space of divergence-free vector fields. From Sec. 2.2, we deduce the optimal incompressible velocity field: $\mathbf{v} \leftarrow G_\sigma \star \mathbf{g}$.

Particular care must be taken when incompressibility is required within a subdomain $\Gamma \subset \Omega$ only, for tracking localised incompressible organs like the heart. This is achieved by defining $p \in H_0^1(\Gamma)$, $p = 0$ on Ω/Γ. Although Gaussian smoothing theoretically preserves the divergence, in practice unconstrained velocities close to Γ boundaries ($\partial\Gamma$) may leak inside the incompressible domain due to the Gaussian convolution, ultimately resulting in volume drifts. Yet, vector derivatives and well-designed Gaussian filters, like Deriche recursive filters, commute. We therefore replace the theoretical "project-and-smooth" strategy by a "smooth-and-project" approach that preserves divergence. To limit numerical instabilities, a smooth domain transition is implemented in a narrow band around Γ by diffusing the pressure field p using heat-transfer equation [6]. Algorithm 1 summarises the main steps of our method, henceforth termed *iLogDemons*, which was implemented using ITK [5] and PETSc.

3 Experiments and Results

3.1 Experiments on Synthetic Datasets

iLogDemons were tested on synthetic datasets with known ground truth. Eight 3D volume-preserving whirl transformations ϕ_α were created with whirl angles $\alpha = 10°$ to $80°$ [13]. Within the whirl domain, the L_2-norm varied from $0.52\,mm$ to $4.78\,mm$ while the Jacobian determinant stayed close to 1 (worse value $|\nabla\phi_{\alpha=80°}| = 1 \pm 0.04$ (mean $\pm$ standard deviation SD)). A 3D isotropic Steady-State Free Precession (SSFP) MRI of the heart, cropped to focus on the heart, was warped with the ϕ_α's and altered with slight Gaussian noise (Fig. 1).

Algorithm 1. iLogDemons: Incompressible LogDemons Registration

Require: Initial stationary velocity field $\mathbf{v}^0$. Usually $\mathbf{v}^0 = \mathbf{0}$ (i.e. $\phi^0 = \mathrm{Id}$).

1: **loop** {*over n until convergence*}
2: Compute update velocity $\delta\mathbf{v}^n$, given $\mathbf{v}^{n-1}$.
3: Fluid-like regularisation: $\delta\mathbf{v}^n \leftarrow G_{\sigma_f} \star \delta\mathbf{v}^n$.
4: Update the correspondence velocity: $\mathbf{v}^n \leftarrow Z(\mathbf{v}^{n-1}, \delta\mathbf{v}^n)$.
5: Diffusion-like regularisation: $\mathbf{v}^n \leftarrow G_{\sigma} \star \mathbf{v}^n$.
6: Solve: $\Delta p = \nabla \cdot \mathbf{v}^n$, $p = 0$ on the incompressible domain boundaries.
7: Project the velocity field: $\mathbf{v}^n \leftarrow \mathbf{v}^n - \nabla p$.
8: **return** $\mathbf{v}$, $\phi = \exp(\mathbf{v})$ and $\phi^{-1} = \exp(-\mathbf{v})$.

The 8 warped images T were registered to the test image R using LogDemons and iLogDemons ($\sigma_x = 1\,mm$, $\sigma^2 = 1\,mm$ and $\sigma_f^2 = 1\,mm$). Relative mean square errors in grey level intensities ($RMSE = \|R - T \circ \phi\|_{L_2} / \|R - T\|_{L_2}$), Jacobian determinant and distance to the true field ϕ_α ($DTF = \|\phi - \phi_\alpha\|_{L_2}$) are reported in Fig. 1. Deformation fields estimated by iLogDemons were almost incompressible. Jacobian determinants were always equal to 1 ± 0.02 independently of the strength of the deformation to recover. Image matching accuracy was not affected by the incompressibility constraint (0.6% decrease). But most importantly, iLogDemons significantly improved the accuracy of the deformations. Mean and SD of DTF were systematically lower (average improvements of 29% and 36% respectively). The larger the deformation, the more significant the improvement. This points out the importance of the choice of the deformation model. Regions with homogeneous grey levels provided little information to accurately estimate the underlying deformation (Fig. 1, yellow arrow). The incompressibility constraint coped with this limitation by ensuring that the estimated deformation was of the same type as the true field. Similar conclusions were drawn on experiments with other parameters ($\sigma_x = 2$, $\sigma^2 = 2$, $\sigma_f^2 = \{0.5, 2\}$).

3.2 Application to Cardiac Deformation Recovery

iLogDemons were then used to estimate the 3D left-ventricular myocardium strain from standard anatomical cine MRI of the heart. Such images have good in-plane and temporal resolutions but large slice thickness, which hampers the accurate estimation of cardiac through-plane motion (Fig. 2). As the volume of the heart is almost constant during the cardiac cycle, incompressible registration is believed to improve the estimation of the cardiac deformation.

Anatomical short axis cine SSFP MRI (cMRI) of a patient with heart failure were acquired with multiple breath-holds (Achieva, Philips Medical System, 30 frames, $1.5\,mm^2$ isotropic in-plane resolution, $10mm$ slice thickness). 3D tagged MRI (tMRI) were also acquired during the same exam (CSPAMM, 23 frames, $0.9mm$ isotropic resolution, tag size $\approx 3mm$). No manual tracking of the tag grids was available since this task is extremely difficult due to the 3D motion. All the images fully covered both ventricles and no slice misalignments were detected. The cMRI were linearly resampled to get isotropic voxel sizes. The tMRI were

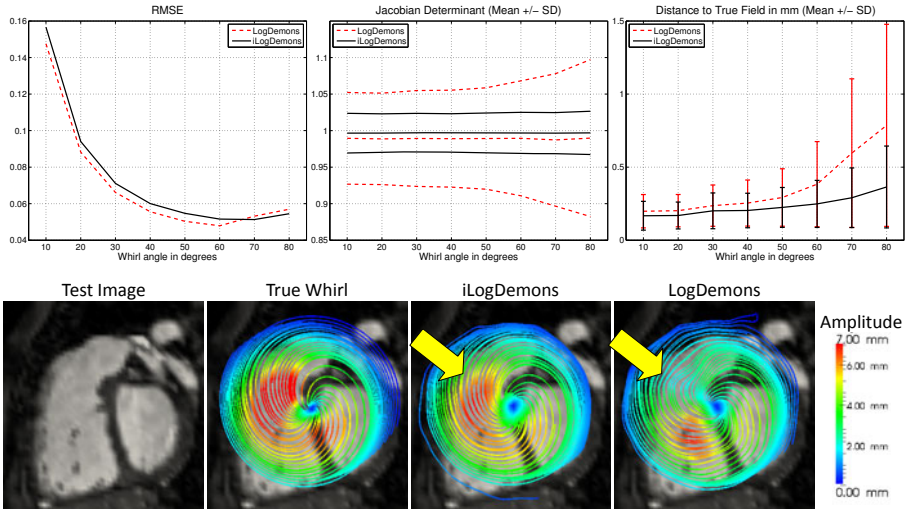


Fig. 1. *Top row*: Results on images warped by 3D synthetic volume-preserving whirls with increasing whirl angle. *Bottom row*: Streamlines of true and estimated whirl deformations (whirl angle $\alpha = 60°$). iLogDemons provided incompressible deformations and outperformed LogDemons in terms of deformation field accuracy (yellow arrow).

spatially and temporally aligned to the cMRI using DICOM information. Because the transformations provided by demons algorithm are resampling fields, myocardium deformation was estimated by recursively registering all the frames of the cardiac sequence to the end-diastole (ED) time frame, as in [9]. Registration parameters were $\sigma_x = 1\,mm$, $\sigma^2 = 2\,mm$, $\sigma_f^2 = 0.5\,mm$ (the smoothing was increased to accommodate the lower image quality). A 2-level multi-resolution scheme was used and registration was stopped as soon as RMSE stopped decreasing. Incompressibility constraint was applied only within the myocardium as volume of surrounding structures like blood pools vary.

First, we estimated the myocardium motion by tracking the heart in the 3D tMRI using iLogDemons. For visual assessment, the deformations were applied to virtual planes manually positioned at ED (Fig. 2, bottom panel). Realistic deformations consistent with the tag grids were obtained, which was further confirmed by the temporal variation of the radial, circumferiential and longitudinal myocardium strains (Fig. 2, green curve). These results were similar to those obtained with logDemons, as the tag grids provided enough texture information in the myocardium to guide the registration. Hence, as no ground truth was available, we considered the iLogDemons estimation as reference. We then estimated the 3D motion of the heart from the cMRI and compared the results with the reference tMRI motion (Fig. 2, blue and red curves). Visually, the warped virtual planes showed that incompressibility constraint did help to recover the longitudinal motion despite the large slice thickness of the cMRI. Estimated longitudinal and circumferential strains confirmed this finding. iLogDemons was closer to the reference than the logDemons (59% of improvement for radial strain, 84% for

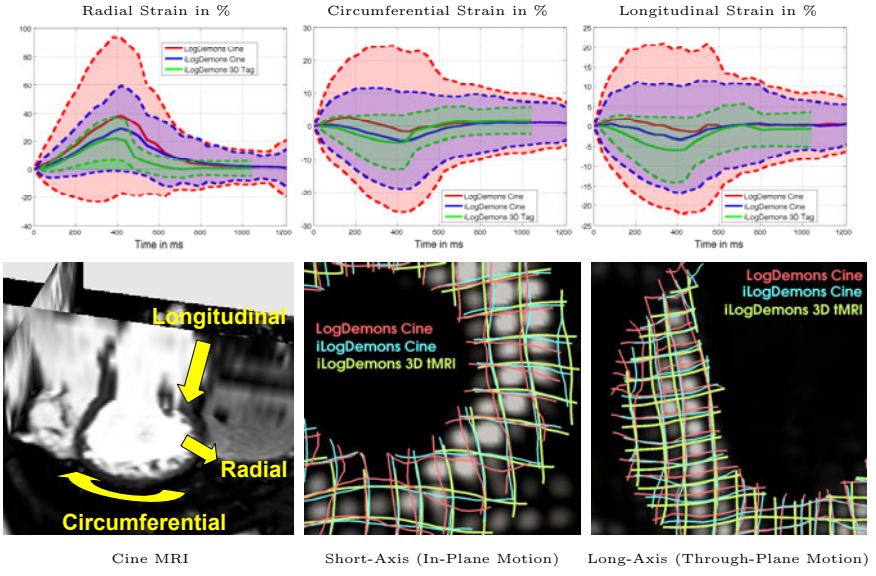


Fig. 2. *Top panels:* Myocardium strains computed from short-axis cine MRI and tMRI. Mean and standard deviation computed over the entire left ventricle. *Bottom panels:* Close-up of the tMRI at end-systole with warped tag planes overlaid. iLogDemons better estimates longitudinal and circumferential motion and strain.

circumferential strain and 42% for longitudinal strain). Radial strain amplitude was more reasonable and the variations of the circumferential and longitudinal strains presented realistic patterns [15], where logDemons estimated an abnormal lengthening at the beginning of the cardiac contraction.

4 Discussion and Future Works

We have adapted logDemons algorithm to provide incompressible deformations. This has been possible by showing that demons Gaussian smoothing minimises an infinite order Tikhonov regulariser. This framework opens the way to new regularisers, such as elastic regularisation. As a result, incompressibility could be ensured by constraining the velocities to be divergence-free. The proposed incompressibility constraint does not introduce any new parameter. Those listed in this paper are present in any recent demons algorithm [14]. One could constrain the correspondence velocity to find the optimal incompressible update deformation. Yet, non-reported experiments showed that this does not significantly improve the results compared to iLogDemons: The updates are usually small and thus near-incompressible. The next step is to modify the demons energy to automatically handle incompressibility in subdomains of the image. From a clinical point of view, we are currently validating this method for the automatic estimation of 3D myocardium strain from standard cardiac images.

Acknowledgements. This work has been partly funded by the European Commission, IST-2004-027749 Health-e-Child Integrated Project. The authors warmly thank the King's College London, St Thomas Hospital, for the cine and tagged MRI.

References

1. 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, p. 924. Springer, Heidelberg (2006)
2. Beg, M., Miller, M., Trouvé, A., Younes, L.: Computing Large Deformation Metric Mappings via Geodesic Flows of Diffeomorphisms. IJCV 61(2), 139–157 (2005)
3. Bistoquet, A., Oshinski, J., Skrinjar, O.: Myocardial deformation recovery from cine MRI using a nearly incompressible biventricular model. Medical Image Analysis 12(1), 69–85 (2008)
4. Bossa, M., Hernandez, M., Olmos, S.: Contributions to 3D diffeomorphic atlas estimation: application to brain images. In: Ayache, N., Ourselin, S., Maeder, A. (eds.) MICCAI 2007, Part I. LNCS, vol. 4791, pp. 667–674. Springer, Heidelberg (2007)
5. Dru, F., Vercauteren, T.: An ITK implementation of the symmetric log-domain diffeomorphic demons algorithm. Insight Journal (May 2009)
6. Evans, L.C.: Partial Differential Equations (June 1998)
7. Ferrant, M., Nabavi, A., Macq, B., Jolesz, F., Kikinis, R., Warfield, S.: Registration of 3D intraoperative MR images of the brain using a finite element biomechanical model. IEEE TMI 20(12), 1384–1397 (2001)
8. Hinkle, J., Fletcher, P., Wang, B., Salter, B., Joshi, S.: 4D map image reconstruction incorporating organ motion. In: Karssemeijer, N., Lelieveldt, B. (eds.) IPMI 2007. LNCS, vol. 4584, p. 687. Springer, Heidelberg (2007)
9. Mansi, T., Peyrat, J.M., Sermesant, M., Delingette, H., Blanc, J., Boudjemline, Y., Ayache, N.: Physically-constrained diffeomorphic demons for the estimation of 3D myocardium strain from cine-MRI. In: Ayache, N., Delingette, H., Sermesant, M. (eds.) FIMH 2009. LNCS, vol. 5528, pp. 201–210. Springer, Heidelberg (2009)
10. Nielsen, M., Florack, L., Deriche, R.: Regularization, scale-space, and edge detection filters. In: JMIV, pp. 291–307. Springer, Heidelberg (1997)
11. Pennec, X., Cachier, P., Ayache, N.: Understanding the "demon's algorithm": 3D non-rigid registration by gradient descent. In: Taylor, C., Colchester, A. (eds.) MICCAI 1999. LNCS, vol. 1679, pp. 597–605. Springer, Heidelberg (1999)
12. Rohlfing, T., Maurer Jr., C., Bluemke, D., Jacobs, M.: Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. IEEE TMI 22(6), 730–741 (2003)
13. Saddi, K.A., Chefd'hotel, C., Cheriet, F.: Large deformation registration of contrast-enhanced images with volume-preserving constraint. In: SPIE Medical Imaging, vol. 6512 (2007)
14. 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)
15. Moore, C., Lugo-Olivieri, C., McVeigh, E., Zerhouni, E.: Three-dimensional systolic strain patterns in the normal human left ventricle: Characterization with tagged MR imaging. Radiology 214(2), 453–466 (2000)

Correspondences Search for Surface-Based Intra-Operative Registration

Thiago R. dos Santos*, Alexander Seitel,
Hans-Peter Meinzer, and Lena Maier-Hein

German Cancer Research Center, Div. Medical and Biological Informatics,
Heidelberg, Germany
t.santos@dkfz.de

Abstract. Intra-operative registration is one of the main challenges related to computer-assisted interventions. One common approach involves matching intra-operatively acquired surfaces (e.g. from a laser range scanner) to pre-operatively acquired planning data. In this paper, we propose a new method for correspondences search between surfaces, which can be used for the computation of an initial alignment. It generates graph representations and establishes correspondences by maximizing a global similarity measure. The method does not rely on landmarks or prominent surface characteristics and is independent on the initial pose of the surfaces relative to each other. According to an evaluation on a set of liver meshes, the method is able to correctly match small submeshes even in this presence of noise.

1 Introduction

Computer-assisted interventions generally require registration of pre-operative planning data with the patient's anatomy. Usually this registration is performed through the use of landmarks. Another approach to achieve this is to acquire intra-operative surfaces (e.g. with a laser range scanner) and match them to surfaces extracted from the planning data. Some commercial systems integrate such surface acquisition technologies for marker-less registration (e.g. BrainLab VectorVision®). However, a complete view of the organ of interest is usually not available in intra-operative situations, making only the acquisition of a partial surface possible, which must be matched to a reference surface [1]. Furthermore, many physical factors, such as lightning and reflexivity, may influence the acquisition devices and generate noise. In general, the process of surface matching comprises three steps: *feature extraction*, *correspondence search* and *transformation computation* [2]. Regardless of the choice of a transformation class (rigid, affine, free-form deformation) and the method for computing it, the matching result depends crucially on the correspondences established in the second step.

Despite the various surface matching methods that have been presented so far (e.g. [3,4]), we found that most of them rely on the identification of points in

* T.R dos Santos is financed by the CAPES/DAAD (Brazil/Germany) scholarship program. The authors would like to thank Tobias Heimann for the helpful insights.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 660–667, 2010.

prominent regions and disregard the spatial relation between descriptors, thus neglecting contextual information that can be crucial for the correct and unambiguous match of small submeshes. In this paper, we present a graph matching-based method for finding correspondences between region features which does not rely on prominent characteristics or on an initial alignment of the meshes. This method is based on the determination of a maximal global similarity, instead of searching for common subgraphs, making it robust to noise.

2 Methods

Our method for correspondences search is performed according to the following procedure (Fig. 1): First, the input meshes are segmented (sec. 2.1). Next, a graph representation is generated from both meshes (sec. 2.2). A graph matching method based on the maximization of a global similarity measure is used for establishing correspondences (sec. 2.3). A method for the elimination of false assignments is employed in the last stage of the graph matching procedure. The established correspondences can then be used to rigidly align the two surfaces, using e.g. the centroids of the assigned region pairs as corresponding points.

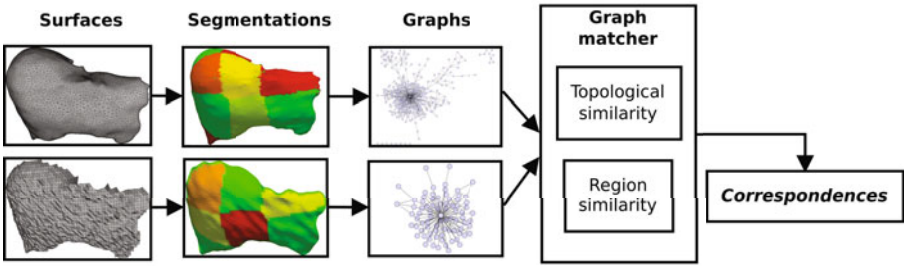


Fig. 1. The correspondence search pipeline. In order to match two surfaces, they are smoothed, segmented and graph representations are created. Those graphs are than matched in order to establish correspondences between the surfaces, by the maximization of the global node assignment similarity.

2.1 Mesh Segmentation

In this step, the input meshes are subdivided into regions that share similar descriptors. In theory, any method for mesh segmentation can be used here. In this paper, we adopted two curvature-based measures that describe the local shape of a particular mesh: the *shape index* and *curvedness*, which are both position and orientation independent [5]. Given two adjacent vertices, v_1 and v_2, the distances in shape index ($\triangle s$) and curvedness ($\triangle c$) between v_1 and v_2 are computed according to the following equations:

$$\triangle s(v_1, v_2) = |s(v_1) - s(v_2)| \tag{1}$$

$$\triangle c(v_1, v_2) = \begin{cases} \frac{\max(c(v_1), c(v_2))}{\min(c(v_1), c(v_2))} - 1 & \min(c(v_1), c(v_2)) > 0 \\ 0 & c(v_1) = 0 \wedge c(v_2) = 0 \\ \infty & otherwise \end{cases} \quad (2)$$

Because of the fact that the curvedness scales inversely with size [5] and is thus dependent on the actual size of the object, computing its distance as a quotient leads to a scaling invariant measure. In the case where one of the vertices is associated with a planar surface region ($c(v) = 0$), there is either no variation, when both vertices are planar, or it is set to ∞ (maximal distance).

Starting at any vertex v, we use local distances of shape index and curvedness between it and its neighbors to segment the mesh. Vertices with descriptor distances smaller than given threshold values (t_s for shape index and t_c for curvedness) are assigned to the same region. Each region R is then represented by a 2-tuple $\bar{R} = (\bar{s_R}, \bar{c_R})$, which is composed of the means of the shape indices and the curvedness values of all points contained in it.

2.2 Graph Construction

Once a surface S has been partitioned into regions, an attributed graph $G_S = (V_S, E_S)$ can be constructed, where V_S is a set of nodes and E_S is a set of arcs. Every $n \in V_S$ represents a surface region R and is attributed with the respective representative $\bar{R}$. Every arc $e \in E_S$ represents a neighborhood relation between two regions R_i and R_j on S. Each arc e is oriented from the region with lower curvedness to the region with higher curvedness.

2.3 Graph Matching

Given two graphs $G_A = (V_A, E_A)$ and $G_B = (V_B, E_B)$, a similarity matrix between their nodes is computed. This matrix is then used to compute an assignment between the nodes, such that the global similarity (the sum of the similarity of all assigned node pairs) is maximal.

Let us define the matrix $S_V = [p_{n_B n_A}]_{|V_B| \times |V_A|}$, which holds the similarity scores between the nodes of G_A and G_B. S_V is composed of two parts

$$S_V \leftarrow S_T + S_R \quad (3)$$

where, $S_T = [q_{n_B n_A}]_{|V_B| \times |V_A|}$ represents the topological similarity scores, and $S_R = [r_{n_B n_A}]_{|V_B| \times |V_A|}$ represents the region similarity scores. Once, the similarities have been computed, S_V is used to determine an optimal global assignment.

Topological similarity scoring. The topological similarity scores are computed locally for each pair of nodes in $(V_B \times V_A)$ through a neighborhood scoring method [6]. According to this method, two nodes are considered similar if their topological neighborhoods are similar. Starting with initial node scores, the similarity between two particular nodes is computed as the sum of the similarity

between their neighbors. This procedure is repeated until a convergence limit is reached. In this way, scores between nodes propagate along to neighboring nodes at each iteration step.

Let us assume two nodes $n_A \in V_A$ and $n_B \in V_B$, and the functions $s(e)$ that returns the source node of a particular arc e, and $t(e)$ that returns the target node. The scores are computed iteratively as follows:

$$q^k_{n_B n_A} \leftarrow \sum_{s(e_B)=n_B, s(e_A)=n_A} (q^{k-1}_{s(e_B)s(e_A)} + q^{k-1}_{t(e_B)t(e_A)}) + \sum_{t(e_B)=n_B, t(e_A)=n_A} (q^{k-1}_{s(e_B)s(e_A)} + q^{k-1}_{t(e_B)t(e_A)}) \tag{4}$$

until $|q^k_{n_B n_A} - q^{k-1}_{n_B n_A}| \leq \varepsilon$ for every n_A and n_B, where ε denotes the convergence limit.

Instead of setting $q^0_{ij} = 1$, for every i and j, as proposed by Zager and Verghese [6] in the case where no previous information about the nodes is available, we use the region similarity scores to initialize it, with cut off thresholds of ∞. Zager and Verghese [6] have proven equation 4 converges independent of the initial values chosen.

Region similarity scoring. Topological similarity by itself would be insufficient for finding corresponding node pairs because many nodes are topologically identical. To address this issue, we also compute the similarity between the regions using their representatives $\bar{R}$ (sec. 2.1). The similarity scores between them are evaluated through a Gaussian kernel, which assigns higher scores to smaller descriptor distances, while giving smaller scores to higher distances. Assuming two region representatives $\bar{R}_1$ and $\bar{R}_2$, we denote their shape index and curvedness distances as $\bar{\triangle}s(\bar{R}_1, \bar{R}_2)$ and $\bar{\triangle}c(\bar{R}_1, \bar{R}_2)$ respectively (sec. 2.1). The Gaussian kernel is defined as follows:

$$g(\triangle, \sigma, \tau) = \begin{cases} -\infty & d > \tau \\ e^{-\frac{\triangle^2}{\sigma^2}} & d \leq \tau \end{cases} \tag{5}$$

where $\triangle$ is the distance, σ is the kernel width and τ is the threshold. Having $n_A \in V_A$ and $n_B \in V_B$, the region similarity matrix is computed as follow:

$$r_{n_B n_A} \leftarrow g(\bar{\triangle}s(\bar{R}_A, \bar{R}_B), \sigma_s, \tau_s) + g(\bar{\triangle}c(\bar{R}_A, \bar{R}_B), \sigma_c, \tau_c) \tag{6}$$

for every n_A and n_B. In the equation σ_s and τ_s denote the shape index kernel width and threshold parameters, and σ_c and τ_c denote the curvedness ones.

Assignment computation. In this stage, an assignment between the nodes of both graphs is computed, such that the global similarity is maximal. This problem is known as assignment problem and there are several methods to solve it [7]. Those methods take a cost matrix and create assignments between each row element to a column element, such that the sum of the costs of the assigned elements is maximized. We use the node similarity scores matrix S_V as a cost matrix and apply the popular *Munkres'* algorithm [8].

Elimination of false assignments. In order to reduce false matches among the assigned nodes, we adopt the premise that, if a particular node $n_A \in V_A$ was assigned to a node $n_B \in V_B$, there must be some other nodes in the neighborhood of n_A that were assigned to nodes in the neighborhood of n_B. If this premise is not confirmed, then the assignment between these nodes is probably incorrect and is removed.

Let N_n^h denote the topological neighborhood with radius h of a particular node n, and $C(N_{n_A}^{h_A}, N_{n_B}^{h_B})$ be a function that returns the number of nodes in the set $N_{n_A}^{h_A}$ that are assigned to nodes in $N_{n_B}^{h_B}$. If n_A was assigned to n_B in the assignment computation stage, this assignment is eliminated if the following equation does not hold:

$$C(N_{n_A}^{h_A}, N_{n_B}^{h_B}) \geq \beta \tag{7}$$

where β denotes the minimal number of node assignments that must exist between $N_{n_A}^{h_A}$ and $N_{n_B}^{h_B}$ in order to maintain the assignment between n_A and n_B.

3 Evaluation

We performed three evaluation studies on five liver meshes, to assess the robustness of the proposed method with respect to the submesh size (sec. 3.1), the influence of the shape descriptor (sec. 3.2), and the influence of noise (sec. 3.3). In all studies, an initial alignment between the shapes was obtained through a transformation that matched the centroids of the assigned regions in a least squares sense.

3.1 Evaluation of Robustness with Respect to the Submesh Size

The following experiment was performed for each reference surface: For each integer i in $[1, 50]$, 500 random submeshes (samples) with an area making up $i\%$ of the area of the reference mesh were generated with a region growing method. Each of the samples was translated and rotated with a random rigid transformation, and the proposed correspondence search algorithm was applied to initially realign the submesh with the reference mesh. Subsequently, ICP was performed to adjust the positioning.

To assess the accuracy of the alignment, the percentage of correctly classified regions and the percentage of incorrectly classified regions was determined. Furthermore, the quality of the final match (after ICP) was evaluated by calculating all distances between the transformed submesh vertices and their corresponding vertices in the reference mesh. The match was considered correct, if all distances were smaller than 10^{-6} (note that all meshes were scaled to fit into the unit box). The parameters were chosen empirically on a set of different meshes as follows: $t_s = 0.3$; $\sigma_s = \sigma_t = 0.1$; $\tau_s = \tau_c = 0.5$; $h_A = h_B = 2$; $\beta = 8$. The curvedness threshold (t_c) was set according to the submesh size to be matched: The smaller the submesh size, the smaller the threshold. This way, a reduction of the size of the

graphs and thus to a reduction of processing time could be achieved for relatively big submeshes. The curvedness threshold was linearly increased from 0.1 to 2.

3.2 Curvature Classes vs. Continuous Curvature Measures

To evaluate the influence of the the shape descriptor on the presented correspondence search method, we repeated the first experiment with an alternative mesh segmentation method. As originally proposed by Besl *et al.* [9], the surfaces were partitioned based on curvature *classes*, which are determined through the sign of the Gaussian and mean curvatures.

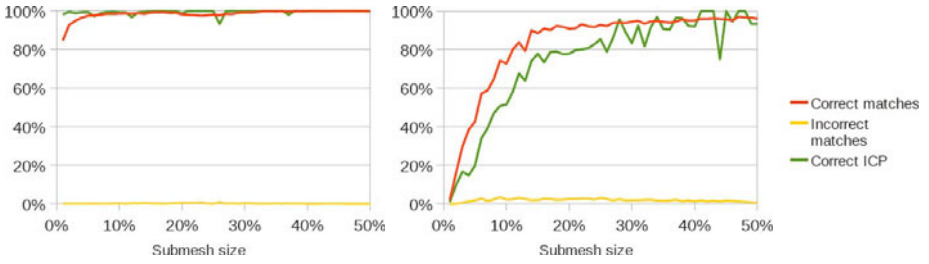
3.3 Evaluation on Noisy Data

In the case of noisy data, the parameters of the segmentation algorithm must be chosen carefully and in an application-specific manner. Two livers were used to empirically optimize the mesh segmentation parameters for each submesh size $i \in \{10\%, 20\%, 30\%, 40\%, 50\%\}$. The smaller the submesh gets, the smaller the tolerance in the mesh segmentation, yielding bigger graphs. Bigger graphs increase the number of nodes that are common to both meshes and help identifying correct correspondences.

The parameters were chosen empirically on a set of different meshes as follows: $\sigma_s = 0.05$; $\sigma_t = 0.15$; t_s =(0.15, 0.15, 0.12, 0.12, 0.1) and t_c =(0.15, 0.13, 0.12, 0.1, 0.03) for submeshes of size (50%, 40%, 30%, 20%, 10%) respectively; $h_A = h_B = 2$; $\beta = 2$. For each submesh size, a set of 50 random submeshes were extracted, and a random vector with direction and magnitude drawn randomly from the interval $[0, 8]$ mm was added to each vertex to simulate noise in the data. The submeshes were then smoothed through the method proposed in [10]. Unlike in the previous experiments, the segmentation of the submesh was potentially different from the segmentation of the corresponding area in the reference mesh due to the noise in the data. Hence, there was generally no isomorphism in the graphs, making an evaluation of the correctly assigned nodes impossible. However, vertex correspondences were known. We assessed the quality of the alignment by computing the percentage of submesh vertices that were assigned to their corresponding ones in the reference mesh after the final iteration of ICP.

4 Results

The results of our evaluation are shown in Figures 2 and 3. When there was partial isomorphisms between the graphs (Fig. 2a), i.e. there was no noise added to the data, the percentage of correct and incorrect node assignments averaged over all five livers ranged from $83.5\% \pm 6.1\%$ and $0.3\% \pm 0.3\%$ respectively (submesh size: 1%) to $97.9\% \pm 3.1\%$ and $0.1\% \pm 0.1\%$ respectively (submesh size: 50%). The correct transformation was found in almost all cases ($94.9\% \pm 11.9\%$ for submesh size: 1%, $100.0\% \pm 0.0\%$ for submesh size: 50%). Depending on the submesh size, processing times for the correspondence search including mesh



(a) Continuous measures (b) Curvature classes

Fig. 2. Results of the surface matching experiment described in sections 3.1 (a) and 3.2 (b). For each reference mesh, the mean percentage of correct matches and incorrect matches after the initial alignment averaged over 500 samples is shown as a function of the submesh size. *Correct ICP* represents the quality of the final match (after ICP) as described in section 3.1. The values are averaged over the five reference surfaces.

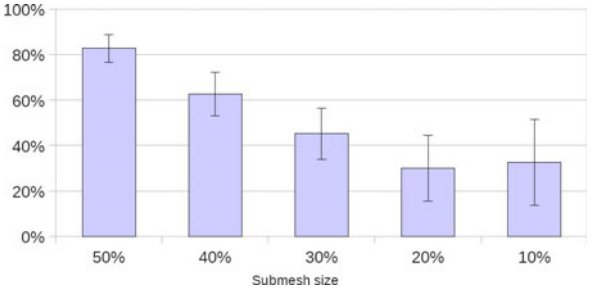


Fig. 3. Results of the experiment described in section 3.3, where the proposed correspondence search method was evaluated with noisy data. The mean percentage (averaged over 50 samples) of submesh vertices that were assigned to their corresponding vertices in the reference mesh, after the final iteration of the ICP, is shown as a function of the size of the submesh. The values are averaged over the five reference surfaces.

segmentation and graph generation ranged from 1 ms to 1.64 s. All processing times were measured using a non-threaded 2.4 GHz Intel machine.

The segmentation based on curvature classes yielded significantly worse results than the segmentation based on shape index and curvedness (Fig. 2b). In the former case, we obtained a percentage of correct and incorrect node assignments ranging from $1.0\% \pm 3.7\%$ and $0.1\% \pm 0.2\%$ respectively (submesh size: 1%) to $94.9\% \pm 1.9\%$ and $0.3\% \pm 0.3\%$ respectively (submesh size: 50%). For small submeshes ($\leq 10\%$) the ICP did not converge into the global optimum. Processing times for correspondence search ranged from 1 to 10 ms.

For noisy data (Fig. 3), the mean percentage of submesh vertices that were assigned to their corresponding vertices in the reference mesh, after the final iteration of the ICP, ranged from $32.5\% \pm 19.3\%$ (submesh size: 10%) to $82.9\% \pm 6.1\%$ (submesh size: 50%). Processing times ranged from 29 to 100 seconds.

5 Discussion

Although surface matching has been subject to considerable research efforts in the past decades, *partial* surface matching with small submeshes remains challenging. This holds especially when no prominent mesh features are available and in the presence of noise. According to our evaluation, the method presented in this paper is highly accurate, yielding a perfect match in almost all cases for submeshes without noise and that made up at least 5% of the size of the reference meshes used in the study. In the presence of noise, good matches were still obtained for submeshes larger than 30% of the mesh size.

We are aware of the fact that we used a relatively simple method for mesh segmentation by combining a classical region growing approach with a surface descriptor that is rather unstable in the presence of noise. However, the good performance of our algorithm despite these conditions demonstrates the potential of the graph-based registration approach. On the other hand, the results obtained with curvature classes indicate that the performance depends considerably on the descriptor and segmentation technique chosen.

Although our approach leaves room for future improvements, we conclude, that (1) our correspondence search algorithm proved to be accurate and robust even for very small submeshes given that these meshes represent exact instances of some reference mesh part and (2) that the results on noisy data are promising and suggest further development of the approach for intra-operative registration purposes.

References

1. Cash, D.M., Miga, M.I.: Compensating for intraoperative soft-tissue deformations using incomplete surface data and finite elements. IEEE Trans. Med. Imag. 24(11), 1479–1491 (2005)
2. Audette, M.A., Ferrie, F.P., Peters, T.M.: An algorithmic overview of surface registration techniques for medical imaging. Med. Img. Anal. 4, 201–217 (1999)
3. Gelfand, N., Mitra, N.J., Guibas, L.J., Pottmann, H.: Robust global registration. In: Proc. EG Symp. Geo. Proc., pp. 197–206 (2005)
4. Gal, R., Cohen-Or, D.: Salient geometric features for partial shape matching and similarity. ACM Trans. Graph. 25(1), 130–150 (2006)
5. Koenderink, J.J., van Doorn, A.J.: Surface shape and curvature scales. Img. & Vis. Comp. 10(8), 557–565 (1992)
6. Zager, L.A., Verghese, G.C.: Graph similarity scoring and matching. App. Math. Let. 21, 86–94 (2008)
7. Dell'Amico, M., Paolo, T.: Algorithms and codes for dense assignment problems: the state of the art. Disc. App. Math. 100, 14–48 (2000)
8. Bourgeois, F., Lassalle, J.C.: An extension of the munkres algorithm for the assignment problem to rectangular matrices. Comm. ACM 14(12), 802–804 (1971)
9. Besl, P.J., Jain, R.C.: Invariant surface characteristics for 3d object recognition in range images. Comp. Vis. Graph. & Img. Proc. 33(1), 33–80 (1986)
10. Jones, T.R., Durand, F., Desbrun, M.: Non-iterative, feature-preserving mesh smoothing. In: SIGGRAPH, pp. 943–949. ACM Press, New York (2003)

Model-Based Multi-view Fusion of Cinematic Flow and Optical Imaging*

Mickael Savinaud[1,2,3], Martin de La Gorce[1], Serge Maitrejean[3], and Nikos Paragios[1,2]

[1] Laboratoire MAS, École Centrale Paris, France
[2] Equipe GALEN, INRIA Saclay - Île de France, Orsay, France
[3] Biospace Lab, Paris, France

Abstract. Bioluminescence imaging (BLI) offers the possibility to study and image biology at molecular scale in small animals with applications in oncology or gene expression studies. Here we present a novel model-based approach to 3D animal tracking from monocular video which allows the quantification of bioluminescence signal on freely moving animals. The 3D animal pose and the illumination are dynamically estimated through minimization of an objective function with constraints on the bioluminescence signal position. Derived from an inverse problem formulation, the objective function enables explicit use of temporal continuity and shading information, while handling important self-occlusions and time-varying illumination. In this model-based framework, we include a constraint on the 3D position of bioluminescence signal to enforce tracking of the biologically produced signal. The minimization is done efficiently using a quasi-Newton method, with a rigorous derivation of the objective function gradient. Promising experimental results demonstrate the potentials of our approach for 3D accurate measurement with freely moving animal.

1 Introduction

Non-invasive visible light imaging is now a widely accepted technology allowing researchers to follow many biological processes in animals [1]. The detection of the light emitted by a probe provides functional information and localization of the processes to be studied. The main limitation of such a modality is the difficulty to localize the signal in 3D especially in bioluminescence imaging techniques (BLI). Indeed photons emitted by bioluminescent cells are strongly scattered in the tissue of the subject and light propagation is diffusive by nature. Therefore different devices and reconstruction methods have been considered to solve this problem [2] but they all require surface acquisition. Furthermore, most of the existing techniques assume that animals have been anesthetized or are immobile, limiting the interest of this modality in functional experiments [3]. However new

* The authors thank Dr. S. Bonzom (Biospace Lab) for his technical assistance and would like to acknowledge the contribution of Dr. R. Boisgard (CEA-LIME, Orsay, France) for *in vivo* experiments.

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 668–675, 2010.

optical imaging devices are now able to image and quantify these processes in freely moving animals in 2D case [4,5].

Prior work trying to tackle these problems includes different techniques of computer vision and various hardware configurations. In Kuo et al. [6], mouse surface topography is obtained using a structured light combined with a single view detector. However this technique does not support freely moving animals because the hardware configuration does not enable cinematic acquisition. The use of temporal information involves either animal tracking or registration of the surface for different poses. In Papademetris et al. [7] a framework that capture articulated movement of the subparts in serial x-ray CT mouse images is proposed. This work has been enhanced to the whole body of the mouse with the use of a skeleton atlas [8] but is restricted to x-ray modality which provides intrinsically 3D information.

Pose estimation and tracking are well known problems in the computer vision community. Discriminative methods aim to recover pose from a single frame through classification or regression techniques [9]. However the high dimensionality of the space spanned by all possible pose restricted these methods to recognition of a limited set of predefined poses. Model-based methods are good candidates for continuous tracking over consecutive frames with small or predictable inter-frame displacements [10,11]. Another interesting aspect of model-based methods is that multi-view data can be handled without solving any correspondence problem between the images. Moreover the matching errors with 2D features on all the cameras can simply be summed to define a single error that should to be minimized.

The aim of this paper is to estimate the animal pose during a cinematic acquisition with a freely moving animal while providing accurate bioluminescence measurement. Our approach is a model-based one where the multi-channel flows are considered as an observation of the scene. In this context an articulated skeleton has been designed to deform the surface mesh towards producing different poses. The estimation of the 3D pose is solved through the optimization of on objective function that aims to generate the observed views from the model while detecting a consistent optical imaging signal across time. We propose a robust derivation of the criteria with respect to scene parameters using a classical gradient optimization.

2 Model Based Articulated Tracking

The proposed approach is inspired from [11] and is extended to the multi-view and multi-channel context. The multi-channel data $I_i = \{V_{i,j}, O_{i,j}\}$, consists of the information obtained by the video acquisition of the moving object $V_{i,j}$ in the different views j as well as the biological data $O_{i,j}$ that are simultaneously recorded on the same views. The goal of our approach is to evaluate the 3D pose with the population of the images by taking advantage of both channels and multi-views. In order to estimate the 3D pose that would correspond to the different observations, the problem will be cast as an energy minimization one.

2.1 Multi-views Pose Estimation

The mouse surface is deformed according to pose changes of an underlying articulated skeleton using Skeleton Subspace Deformation (SSD) [12,13]. The skeleton comprises 20 bones with 64 degrees of freedom (DOF). Each DOF corresponds to an articulation angle whose range is bounded to avoid unrealistic poses of the mouse. The mouse pose is fully determined by a vector $\mathbf{\Theta} = [\mathbf{w}, \mathbf{t}, \mathbf{q}]$ that comprises 57 articulation parameters vector $\mathbf{w}$, the 3D translation vector $\mathbf{t}$ and a quaternion $\mathbf{q}$ that specifies the global position and orientation of the mouse body with respect to the world's coordinate frame. In order to adapt the size of the mouse, three additional morphological parameters are introduced for each bone. These scale factors are added to the $\mathbf{\Theta}$ parameters that are optimized while fitting the model to the observations in the first frame and are kept constant for the subsequent frames.

The lighting is modeled as four point sources placed at an infinite distance and an ambient light. It is parameterized using three directional components for each light and with an additional ambient component, which produces a vector $\mathbf{L}$ of 13 parameters. The complexity of the lighting conditions is enforced by the fact that in our experiments light sources produce localized light spots due to high directivity of the light at output of the optical fibers. The mouse skin surface is assumed to be Lambertian. The mouse is white and we can assume the albedo to be constant over its entire surface. Thus we do not require the use of a texture mapped onto the surface due to the small variations of the albedo.

For a given pose $\mathbf{\Theta}$ and an illuminant $\mathbf{L}$, we define $V_{\text{syn},j}(\mathbf{x}; \mathbf{\Theta}, \mathbf{L})$ to be the RGB intensities of the corresponding synthetic image comprising the mouse and the background evaluated at the point location $\mathbf{x}$ from the j^{th} camera. This is formulated using a classical perspective projection, the hidden surface removal and the Gouraud shading model. The tracking process attempts to recover for each successive frame the pose parameters $\mathbf{\Theta}$ and the illuminant $\mathbf{L}$ that produce the three synthesized images that best match the three observed ones, denoted by $V_{\text{obs},j}$, with $j = 1, \ldots, 3$ the index of the camera. In the following objective function:

$$E_V(\mathbf{\Theta}, \mathbf{L}) = \sum_{j=1}^{3} \int_{\Omega} \underbrace{\rho\big(V_{\text{syn},j}(\mathbf{x}; \mathbf{\Theta}, \mathbf{L}) - V_{\text{obs},j}(\mathbf{x})\big)}_{R_j(\mathbf{x}; \mathbf{\Theta}, \mathbf{L})} d\mathbf{x}, \tag{1}$$

the main term is defined by summing the residual errors $R_j(\mathbf{x}; \mathbf{\Theta}, \mathbf{L})$ between the synthetic images and the observed images V for each of the three cameras.

2.2 Bioluminescence Position Constraints

In order to take advantage of the information provided by the BL images $O_{\text{obs},i,j}$, we compute in the first image the 3D position of the bioluminescence by automatic detection of the BL spot in each view. The 3D position X^O_{obs} of the light source can be estimated using a standard triangulation method. We do not adopt complex bioluminescence tomography methods because the tumors position is not expected to be far from the mouse surface. In case of tumors, we assume

that this point is rigidly fixed to its nearest bone in the model. For each frame i, we detect automatically the position of the bioluminescence spot $P^{O}_{obs,i,j}$ in each view if possible. We aim at minimizing the sum of retroprojection error between these points and X^{O}_{obs}. We are now able to compute the expected position of X^{O}_{obs} given any new candidate mouse pose parameter vector $\mathbf{\Theta}$.

$$E_O(\mathbf{\Theta}) = \sum_{j=1}^{3} \|\Pi_j(X^{O}_{\text{obs}}(\mathbf{\Theta})) - P^{O}_{\text{obs},i,j}\|^2 \tag{2}$$

where Π_j corresponds to the operation of 3D to 2D projection using the j^{th} BL detector. E_O sums over the three views the 2D distances between the projection of the predicted bioluminescence source position X^{O}_{obs} and the actual observation extracted in the new $O_{i,j}$ image. This new term enforces the pose estimation of the mouse with respect to the BL signal during the tracking and enables to exploit in minimization process the biological information provided by secondary camera.

2.3 Tracking with Energy Minimization

During the tracking we determine, for each frame, the pose and the illumination parameters by minimizing an objective function which combines the two previous formulas. A factor β weights the two energies and is chosen empirically to be the squared inverse of the maximum expected deviation between the observed signal and the one fixed to the model. The minimization is done efficiently using a quasi-Newton method that requires the gradient of the objective function E_V. The gradient with respect to the lighting parameters is obtained by using the differentiation chain rule on the residual intensities. The gradient with respect to the pose $\mathbf{\Theta}$ is not straightforward to derive due to discontinuities in the residual image along the occlusion boundaries when $\mathbf{\Theta}$ varies. The adequate treatment of these discontinuities when computing the gradient is done using the proposed occlusion forces in [11].

3 Experimental Validation

Experiments were conducted using an innovative device capable of recording simultaneously scene video and optical data at 43 *fps*. The scene video V is acquired under near IR lighting and the BL signal O is recorded by an intensified CCD (Photon Imager, Biospace Lab). The two signals are simultaneously recorded and spatially registered [4]. Towards acquiring simultaneously different views of the animal and the BL signal emitted without large hardware modifications, we have considered two mirrors. Mirrors are defined by planes which are placed on the device stage with a angle of 90 degrees somewhere in the V camera field of view. The image of the mouse seen in each mirror can be interpreted as the image seen from a virtual camera, whose position and orientation are obtained by reflection with respect to the corresponding mirror plan (Fig. 1-C).

The parameters of the cameras are determined using the calibration toolbox. Mirror parameters are manually optimized with a known object to provide virtual camera positions and orientations. Illumination is provided by four optical fibers placed at the top of the the scene and at each extremity of the scene. The mouse can move in an area of 5 cm by 18 cm.

The mouse model used for the pose estimation is composed of a skeleton of 20 bones manually segmented from static micro-CT acquisitions (Skyscan 1178, Skyscan) and guided by a anatomical book [14] (Fig. 1-A). The mouse surface is modeled as a three dimensional, closed and orientable triangulated surface of 1252 facets (Fig. 1-B). The mesh of the mouse was created with the micro-CT surface and elements of computer graphic project on mouse animation. The extremities of legs have not been modeled because it appeared through experiment that tracking these parts of the mouse is difficult given the quality of our observations while not being useful for our application (tumor cells embedded on the top of the mouse).

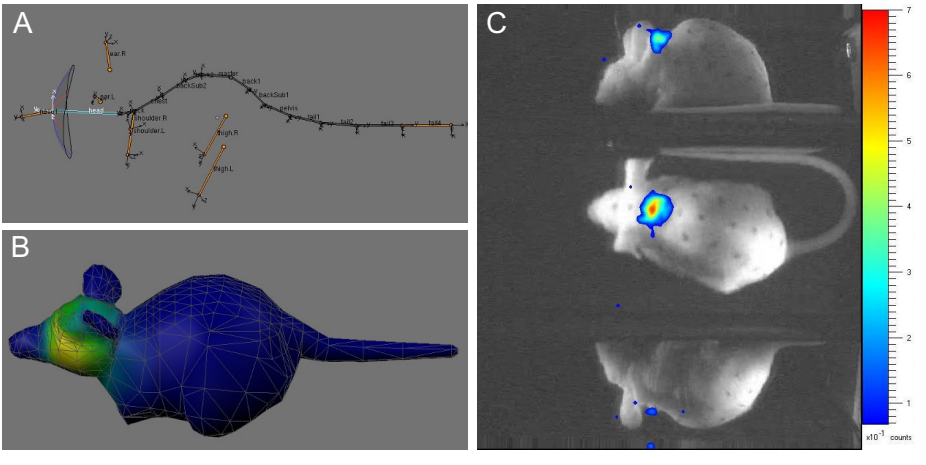


Fig. 1. Model and observations. On the left the skinned mouse model. On the right, fusion of the observed video and bioluminescence signal in the multi-view device.

This framework was applied to image a freely moving mouse (NMRI) bearing a PC12 tumor injected ten days before experiments in the dorsal part of neck (10000 cells in 0.5μL). In addition, we have drawn onto the surface of the mouse landmarks to measure locally the 3D position of the mouse surface. To validate our approach, we tested our method on 4 acquisitions which represent a total of 580 frames. Visual assessment and 3D cinematic analysis of the bioluminescence signal are used to demonstrate the interest of the method for measurement on freely moving animal.

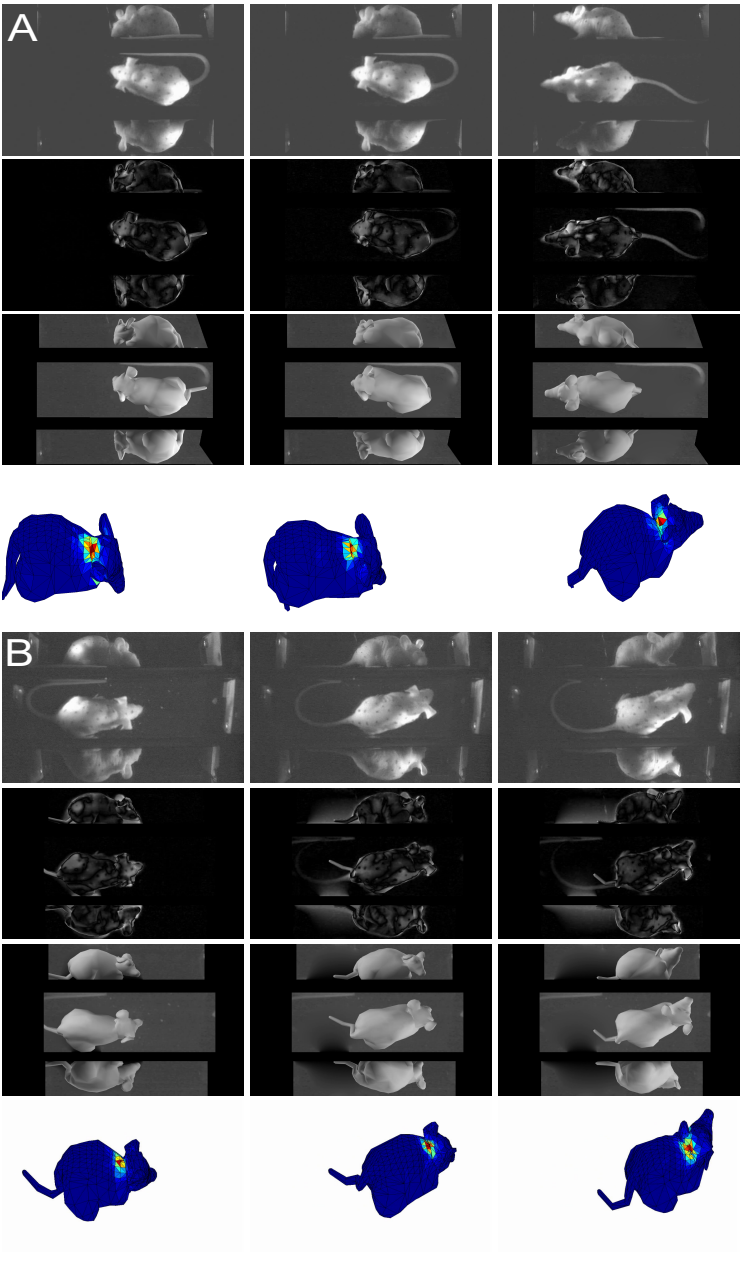


Fig. 2. Two sequences (A and B) processed with our tracking method. Each row corresponds to: the observed image, the final residual image, the synthetic final image and the model with the final pose and the bioluminescence backprojected to the surface.

3.1 Comparison with Triangulated Data

In order to evaluate our estimation of the pose, we manually annotated throughout the first sequence 800 couples of points in two different views. For each of these couples, we assigned a 3D position on the mesh with its first annotation. In the other frames, the new position of this point on the mesh was estimated with a triangulation method. Computing a 3D distance between the position of the reference with the corresponding pose provides a way to estimate the error produced by our method. In the first sequence, this error was about 5 mm with our 800 visual correspondences.

3.2 Visual Assessment

In order to check the usability of our pose estimation, we studied the biological signal throughout a sequence using visual assessment. Fig. 2 shows 6 frames extracted from the videos provided as additional material (respectively sequence 1 and 3). The low residual implies the correspondence between synthetic data and observations. Residual artifact are due to difficulty to render light spots generated by optical fibers with our illumination model. The backprojection of the BL signal on the surface is computed with all the views recorded by the camera O with a temporal smoothing of 5 frames and a spatial smoothing of 3 mm. In the great majority of frames the signal of interest is registered to the corresponding place of the emission surface. To our knowledge, this type of measurement is compatible with optical imaging experiments.

3.3 3D Cinematic Analysis

To evaluate the possibility to perform studies on freely moving animals with this new tool we computed for the first frame a region of interest (ROI) based on the faces which corresponds to the tumors position. In each following frame we compared the signal measured on these faces with the reference one to evaluate the stability and robustness of the pose estimation regarding to the biological data. Along our 4 sequences more than 75% of the signal was kept on the right faces (Table 1).

Table 1. ROI tracking: the two first lines indicate the characteristics of the first ROI while the last evaluates the quantity of the signal following the ROI throughout the sequence

	SEQ 1	SEQ 2	SEQ 3	SEQ 4
Number of faces:	61	41	42	49
Size of ROI (cm^2):	1.83	1.27	1.60	1.65
Mean of ROI intensity similarity:	88%	81%	83%	75%

4 Discussion

In this paper we have proposed a novel approach for multi-view fusion of cinematic flow and optical images of mice. The method explores an analysis-by-synthesis approach where a model involving articulations, surface properties and appearance properties is optimized with respect to the different views. Such optimization is done jointly on the visual/optical image space through the certain constancy hypothesis on the bioluminescence imaging. Promising results demonstrate the ability of the method to deal with freely moving animals and enhance the optical imaging signal towards improved preclinical exploitation. Future work consists of introducing explicit modeling of the bioluminescence sources, and a continuous manner on incorporating constancy on the optical imaging space.

References

1. Weissleder, R.: Scaling down imaging: molecular mapping of cancer in mice. Nature Reviews Cancer 2, 11–18 (2002)
2. Gibson, A.P., Hebden, J.C., Arridge, S.R.: Recent advances in diffuse optical imaging. Physics in Medicine and Biology 50(4), R1–R43 (2005)
3. Hildebrandt, I.J., Su, H., Weber, W.A.: Anesthesia and other considerations for in vivo imaging of small animals. ILAR Journal 49(1), 17–26 (2008)
4. Roncali, E., Savinaud, M., Levrey, O., Rogers, K.L., Maitrejean, S., Tavitian, B.: A new device for real time bioluminescence imaging in moving rodents. Journal of Biomedical Imaging 13(5), 054035 (2008)
5. Rogers, K.L., Picaud, S., Roncali, E., Boisgard, R., Colasante, C., Stinnakre, J., Tavitian, B., Brulet, P.: Non-invasive in vivo imaging of calcium signaling in mice. In: PLoS ONE (October 2007)
6. Kuo, C., Coquoz, O., Troy, T.L., Xu, H., Rice, B.W.: Three-dimensional reconstruction of in vivo bioluminescent sources based on multispectral imaging. Journal of Biomedical Optics 12(2), 024007 (2007)
7. 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)
8. Baiker, M., Milles, J., Vossepoel, A., Que, I., Kaijzel, E., Lowik, C., Reiber, J., Dijkstra, J., Lelieveldt, B.: Fully automated whole-body registration in mice using articulated skeleton atlas. In: IEEE ISBI 2007, pp. 728–731 (April 2007)
9. Favreau, L., Reveret, L., Depraz, C., Cani, M.P.: Animal gaits from video. In: ACM SIGGRAPH Symposium on Computer Animation (2004)
10. Gall, J., Stoll, C., de Aguiar, E., Theobalt, C., Rosenhahn, B., Seidel, H.P.: Motion capture using joint skeleton tracking and surface estimation. In: IEEE CVPR 2009, pp. 1746–1753 (June 2009)
11. de LaGorce, M., Paragios, N., Fleet, D.: Model-based hand tracking with texture, shading and self-occlusions. In: IEEE CVPR 2008, pp. 1–8 (June 2008)
12. Magnenat-Thalmann, N., Laperrière, R., Thalmann, D.: Joint-dependent local deformations for hand animation and object grasping, pp. 26–33 (1988)
13. Lewis, J.P., Cordner, M., Fong, N.: Pose space deformation: a unified approach to shape interpolation and skeleton-driven deformation. In: ACM SIGGRAPH, pp. 165–172 (2000)
14. Cook, M.J.: The Anatomy of the Laboratory Mouse. Elsevier, Amsterdam (1965)

Simultaneous Geometric - Iconic Registration

Aristeidis Sotiras[1,2], Yangming Ou[3],
Ben Glocker[4], Christos Davatzikos[3], and Nikos Paragios[1,2]

[1] Laboratoire MAS, Ecole Centrale de Paris, France
[2] Equipe GALEN, INRIA Saclay - Île de France, France
[3] Section of Biomedical Image Analysis (SBIA), University of Pennsylvania, USA
[4] Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany
aristeidis.sotiras@ecp.fr

Abstract. In this paper, we introduce a novel approach to bridge the gap between the landmark-based and the iconic-based voxel-wise registration methods. The registration problem is formulated with the use of Markov Random Field theory resulting in a discrete objective function consisting of thee parts. The first part of the energy accounts for the iconic-based volumetric registration problem while the second one for establishing geometrically meaningful correspondences by optimizing over a set of automatically generated mutually salient candidate pairs of points. The last part of the energy penalizes locally the difference between the dense deformation field due to the iconic-based registration and the implied displacements due to the obtained correspondences. Promising results in real MR brain data demonstrate the potentials of our approach.

1 Introduction

Image registration is a fundamental problem in medical image analysis. Due to its importance, great efforts have been made to tackle this problem resulting in numerous approaches. The existing methods fit mainly into two categories.

In the first class of methods, *geometric* (e.g. [1,2,3]), landmarks are detected and subsequently matched to establish correspondences between the images being registered. Such an approach exhibits strength and limitations. On one hand, if landmarks are appropriate determined, solving the registration problem is straightforward and the method is not sensitive to the initial conditions and can cope with important deformations. On the opposite side, the registration result is usually accurate in the vicinity of the interest points and its accuracy deteriorates away from them while often the extraction of landmarks is also problematic.

The second class of methods, *iconic* (e.g.[4,5,6,7]), takes advantage of the intensity information of all positions and tries to recover the deformation that optimizes a criterion based on it. Methods of this class exhibit globally better accuracy at the cost of greater computational effort. These approaches consider all points contributing equally to the objective function, thus discarding the importance of the salient points of the image. Furthermore, they are very sensitive to the initial conditions and often unable to deal with large deformations. Last but not least, their performance deteriorates when

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 676–683, 2010.

considering multi-modal image fusion where defining an appropriate similarity metric is far from being trivial, while at the same time the optimization of the objective function becomes more challenging.

During the last years, efforts have been made to bridge the two main classes of approaches by taking advantage of both complementing types of information, resulting in a *hybrid* approach (e.g.[8,9,10,11,12,13]). Among these methods, the way the two types of information are used varies widely. Most methods decompose the registration problem in two separate steps, each one exploiting information of one type. Typically, landmark information is used either to provide a coarse registration that is subsequently refined by using the intensity information [8] or more often, after having established point correspondences, use them in the objective function to ensure that the optimal deformation field will comply with them [10,11,12,13]. However, by considering geometric and iconic information in a non-coupled way, the solution of each subproblem (point correspondence, dense registration) cannot fully profit from the solution of the other. We refer to [9], where an objective function is proposed that combines an iconic-based dense field and geometric constraints and is optimized by alternating between three steps: estimate the dense deformation field, the point correspondences and regularize.

In this paper we couple the point correspondence and the dense registration problem into an unified objective function where the two problems are solved simultaneously in an one step optimization. We employ a Markov Random Field formulation towards introducing individual costs for the family of parameters and their interactions. Moreover, due to the discrete nature of the proposed objective function any intensity-based criterion can be used. To the best of our knowledge, only one other hybrid method can claim that [10], but it cannot guarantee the convergence as a ping-pong effect is possible between iterations. Moreover the proposed method is able to constrain the recovered dense deformation field to be diffeomorphic contrary to the rest of the hybrid methods. Last but not least, the influence of the landmarks is done in a local way and in respect to the deformation model used without the use of heuristics and any assumption on the number or the nature of the landmarks.

2 Iconic-Landmark Registration

Let us consider two images $I_1 : \Omega_1 \mapsto \mathbb{R}, I_2 : \Omega_2 \mapsto \mathbb{R}$, a set of points of interest $\mathcal{P}_1 \in \Omega_1$, and a set of potential candidates for the points $p_1 \in \mathcal{P}_1,\ \mathcal{P}_2 \in \Omega_2$ such that $|\mathcal{P}_2| > |\mathcal{P}_1|$. By $|\cdot|$ we denote the cardinality of the set. The aim of the algorithm is to estimate the deformation field $\mathcal{T} : \Omega_1 \mapsto \mathbb{R}^3$, such that an iconic criterion defined in the whole image domain is satisfied, and to recover the correspondences between the two different point sets such that the estimated solutions are consistent with each other.

A grid-based deformation model is going to be used, resulting in a decreased number of variables to be estimated. The dense deformation field is going to be given by interpolating the displacements of the grid points. Let us consider a deformation grid $G : [1, S_x] \times [1, S_y] \times [1, S_z]$ then,

$$\mathcal{T}(\mathbf{x}) = \mathbf{x} + \mathcal{D}(\mathbf{x}) \text{ where } \mathcal{D}(\mathbf{x}) = \sum_{p \in G} \eta(\|\mathbf{x} - \mathbf{p}\|)\mathbf{d_p}. \tag{1}$$

$\eta(\cdot)$ is a weighting function that measures the contribution of each control point to the displacement field $\mathcal{D}(\cdot)$. In the case of the cubic B-spline that are going to be used here, the weights are given by the cubic B-spline basis functions depending on the distance of the voxel from the control point. The specific deformation model allows diffeomorphic deformations to be guaranteed through the use of hard constraints [5].

The goal will be reached by coupling the dense deformation field estimation and the point correspondence problem in one with the use of the Markov Random Field (MRF) theory. The typical first-order MRF energy is of the form:

$$E_{MRF} = \sum_{p \in \mathcal{G}} V_p(l_p) + \sum_{p \in \mathcal{G}} \sum_{q \in \mathcal{N}(p)} V_{pq}(l_p, l_q) \tag{2}$$

where the first term (unary potentials) encodes the information stemming from the observed variables (intensity values) and typically acts as the data term of the energy. The second term (pairwise potentials) encodes relations between pairs of latent variables and typically acts as a regularizer. By l_p the label attributed to variable p is denoted.

2.1 Point Correspondence Problem

For the point correspondence part, the goal is to estimate which point $p_2 \in \mathcal{P}_2$ corresponds to each of the points $p_1 \in \mathcal{P}_1$. We are assuming that the true underlying anatomical correspondence is included in the set of potential candidates $\mathcal{P}_2$. The two point sets should capture the important geometric information of the two images in order to act as the additional constraints that will enhance the performance of the registration.

Any method for establishing candidate correspondences can be used. Herein, multi-scale and multi-orientation Gabor filters are used to locate points of interest in the image domain. Gabor filters are able to provide distinctive description for voxels belonging to different anatomical regions by capturing local texture information [14,15]. Local texture information reflects the underlying geometric and anatomical characteristics. Thus, points exhibiting a high response to Gabor filters are most likely placed in salient anatomical regions whose matching can be used to guide the registration process. In other words, the set $\mathcal{P}_1$ consists of points whose response to the Gabor filters is significant and that are distributed in space. Then, the set $\mathcal{P}_2$ of potential correspondences can be populated by taking for every $p_1 \in \mathcal{P}_1$, the top K candidate points in an appropriately defined sphere in terms of a similarity criterion that is based on the difference between D-dimensional Gabor attribute vectors $A(\cdot)$ weighted by the mutual saliency. The role of the mutual saliency, $ms(p_1, p_2) = \frac{mean_{n \in s_{in}}(sim(A(p_1), A(n)))}{mean_{n \in s_{out}}(sim(A(p_2), A(n)))}$ [15] (s_{in} and s_{out} are appropriately defined regions around the points and are adaptive to the scale from which Gabor attributes are extracted), is to narrow down the selection to candidate points that are mutually salient indicating matching reliability. The similarity is given by

$$sim(p_1, p_2) = \frac{1}{1 + \frac{1}{D}\|A(p_1) - A(p_2)\|^2}. \tag{3}$$

In a MRF framework, the point correspondence problem can be solved by minimizing an appropriately defined labeling energy. What we search is which label (or index

of candidate point) to attribute to each $p_1 \in \mathcal{P}_1$ to establish a correspondence. Thus, the label set is defined as $\mathcal{L}_{gm} = \{l^1, \cdots, l^K\}$, where the label assignment l_i^j corresponding the j-th potential candidate point to the i-th point/node. The optimal labeling $\mathbf{l}^\star = (l_1, \cdots, l_N)$ will minimize the discrete objective function. For that purpose, we are going to construct a graph $\mathcal{G}_{gm} = (\mathcal{V}_{gm}, \mathcal{E}_{gm})$ where the set of the nodes $\mathcal{V}_{gm}$ coincides with the point set $\mathcal{P}_1$ and each edge in $\mathcal{E}_{gm}$ encodes a geometric compatibility constraint.

The discrete objective function is of the type of Eq.2. The unary potentials will quantify the level of similarity between the landmark and its assigned candidate point.

$$V_p(l_p) = exp\left(-\frac{ms(p,p') \cdot sim(p,p')}{2\sigma^2}\right) \tag{4}$$

where p' is the point in $\mathcal{P}_2$ that is corresponded to p through a label assignment l_p and σ is a scaling parameter.

The regularization term will impose a geometric consistency on the established correspondences. What we would expect from the recovered pairs is that the distance between adjacent pairs should be preserved by their corresponded ones, thus avoiding having landmarks flipping positions. The pairwise potential is defined as:

$$V_{pq}(l_p, l_q) = \|(\mathbf{p} - \mathbf{q}) - (\mathbf{p}' - \mathbf{q}')\| \tag{5}$$

where in bold the physical position of the point is denoted. We consider that an affine registration step has preceded, as a consequence no normalization is needed.

2.2 Iconic Registration

For the estimation of the dense deformation field we follow the approach proposed in [6]. The reasons behind this choice lie in the fact that due to the discrete nature of the formulation a wide range of similarity measures can be used. Moreover, the method is computational efficient while producing precise results. For completeness reasons, the iconic registration method is going to be presented briefly in this section.

Given the deformation model, we aim at optimizing the displacements of the grid points. In the proposed discrete framework this is equivalent to assign a label to each grid node such that the displacement associated to it decreases the energy. For the iconic registration part, the label set $\mathcal{L}_{ic}$ is quantized version of the deformation space where each label l corresponds to a displacement $\mathbf{d}^l$. To impose the diffeomorphic property, the maximum displacement, to which a label is corresponded, is $0.4 \times \delta$ where δ is the grid spacing [5]. In order to solve the optimization problem, a regular graph $\mathcal{G}_{ig} = (\mathcal{V}_{ic}, \mathcal{E}_{ic})$ is going to be constructed. Its nodes coincide with the nodes of the deformation grid G and edges exist between neighboring nodes assuming a 6-connectivity scheme.

The unary potentials are defined as follows:

$$V_p(l_p) = \int \hat{\eta}(\|\mathbf{x} - \mathbf{p}\|)\rho(I_1(\mathbf{x} + \mathbf{d}^{l_p}), I_2(\mathbf{x}))d\mathbf{x} \tag{6}$$

The data term is based on an iconic similarity measure $\rho(\cdot)$ and $\hat{\eta}(\cdot)$ is a function that determines how much a voxel $\mathbf{x}$ influences a node p. It is defined as

$$\hat{\eta}(\|\mathbf{x}-\mathbf{p}\|) = \frac{\eta(\|\mathbf{x}-\mathbf{p}\|)}{\int_{\Omega}\eta(\|\mathbf{y}-\mathbf{p}\|)d\mathbf{y}} \text{ or } \hat{\eta}(\|\mathbf{x}-\mathbf{p}\|) = \begin{cases} 1, & \text{if } \eta(\|\mathbf{x}-\mathbf{p}\|) > 0, \\ 0, & \text{else.} \end{cases} \tag{7}$$

for the case of voxel-wise and more sophisticated statistical criteria respectively. The regularization term in the simplest case can be a vector difference between the displacements that are encoded by the two different labels normalized by the difference of the control points,

$$V_{pq}(l_p, l_q) = \frac{\|\mathbf{d}^{l_p} - \mathbf{d}^{l_q}\|}{\|\mathbf{p}-\mathbf{q}\|} \tag{8}$$

2.3 Simultaneous Geometric - Iconic Registration

In order to tackle both problems at the same time, a new graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ should be considered. The node system of the new graph is simply the union of the nodes of the subproblems $\mathcal{V} = \mathcal{V}_{ic} \cup \mathcal{V}_{gm}$. The edge system of the graph will comprise of the edges of each subproblem and appropriate edges that will connect the two graphs and will encode the consistency between the solutions of the two subproblems $\mathcal{E} = \mathcal{E}_{ic} \cup \mathcal{E}_{gm} \cup \mathcal{E}_{cn}$.

The unary potentials and the pairwise potentials will be the same as the ones previously detailed except from the ones that correspond to the new edges and have yet to be detailed. In order to impose consistency upon the solutions of the two subproblems, the difference between the displacement field due to the grid-based deformation model and the displacement implied by the recovered correspondence should be minimal at the landmark position. Given a cubic B-spline FFD deformation model and considering, without loss of generality, only one landmark, then

$$0 = \|\mathcal{D}(\mathbf{x}_*) - (\mathbf{p}'_* - \mathbf{p}_*)\| = \|\sum_{i=1}^{M}\beta_i(\mathbf{x}_*)\mathbf{d}_{p_i} - \mathbf{d}_{p_*}\| =$$
$$\|\sum_{i=1}^{M}\beta_i(\mathbf{x}_*)\mathbf{d}_{p_i} - \sum_{i=1}^{M}\beta_i(\mathbf{x}_*)\mathbf{d}_{p_*}\| \leq \sum_{i=1}^{M}\beta_i(\mathbf{x}_*)\|\mathbf{d}_{p_i} - \mathbf{d}_{p_*}\| \tag{9}$$

where the displacement of the voxel $\mathbf{x}_*$ is $\mathbf{d}_{p_*} = \mathbf{p}'_* - \mathbf{p}_*$ and the properties of the cubic B-spline, $\sum_i^M \beta_i(\cdot) = 1,\ \beta_i(\cdot) \geq 0$, and the triangular inequality were used. $M = 4 \times 4 \times 4$, the number of the grid nodes that control the displacement of a voxel.

The previous relation (Eq.9) can be modeled by adding edges between the nodes of the irregular grid $\mathcal{V}_{gm}$ and those nodes of the regular grid $\mathcal{V}_{ic}$ that control the displacement of the position in which the landmark is placed. The pairwise potentials are given by the following equation

$$V_{pq}(l_p, l_q) = w\|\mathbf{d}^{l_q} - (\mathbf{p}' - \mathbf{p})\|. \tag{10}$$

p, p' and $\mathbf{d}^{l_q}$ are defined as previous. w is a weight based on the cubic B-spline basis as a function of the distance of the landmark from the control point q. This formulation

results in minimizing an upper bound of the energy while permitting us to model the problem by pairwise relations between the latent variable and thus allowing for the use of any standard MRF optimization technique.

3 Experimental Validation

3.1 MRF Optimization

In order to optimize the resulting MRF problem, the convergent Tree-Reweighted (TRW) Message Passing algorithm was used [16]. The TRW algorithm aims to address the MRF optimization problem by tackling the simpler dual of its Linear Programming (LP) relaxation. Any solution of this problem is a lower bound to the energy of the original problem. Thus, the TRW algorithm aims to maximize the bound and thus reach a solution. TRW is known to have a state of the art performance among the various discrete optimization methods and has proven its applicability in various tasks in the fields of computer vision and medical imaging.

3.2 Experimental Results

To validate the proposed method, a data set of 11 $T1$-weighted brain images of different subjects was used. The resolution of the images is $256 \times 256 \times 181$ with an isotropic voxel spacing of $1mm$. The volumes were manually annotated into 11 classes (background, cerebrospinal fluid, white matter, gray matter, fat, muscle, muscle/skin, skull, around fat, dura matter and bone marrow). The Sum of Absolute Differences (SAD) was used as iconic similarity criterion.

To visually assess the quality of the registration, a template image is chosen and all the rest are registered to it. Then the mean image, its difference with the template as well as the standard deviation image are calculated for the images before and after the registration (Fig.1). The blur mean image and the great values for the standard deviation before registration depict the difficulty of the registration task. On the bottom row, the mean image has become sharper indicating that the images have been spatially normalized. The values of the standard deviation have decreased especially in the area of the ventricles. The results of the registration can be also seen when comparing the two difference images and noting that the difference image (c2) is darker than (c1).

To further quantify the performance of the algorithm, we performed all possible 110 pair-wise registrations and the provided voxel-wise manual segmentations are used to measure the accuracy of the proposed method. We select each one of the images as target and register the rest to it. The obtained deformation fields are then used to warp the segmentations. Finally, we compare the deformed segmentations with the manual ones by calculating the DICE overlap measure, its sensitivity and specificity. The results are presented graphically in Fig.2 in the form of box plots. When comparing the obtained results to the initial DICE coefficients, it becomes evident the impact of the registration. Moreover, we compare the proposed method with the one that only uses the iconic information to show the added value of the use of the landmark information. The iconic registration is performed by discarding the landmark information in the proposed framework. In a similar way, DICE, sensitivity and specificity are computed for

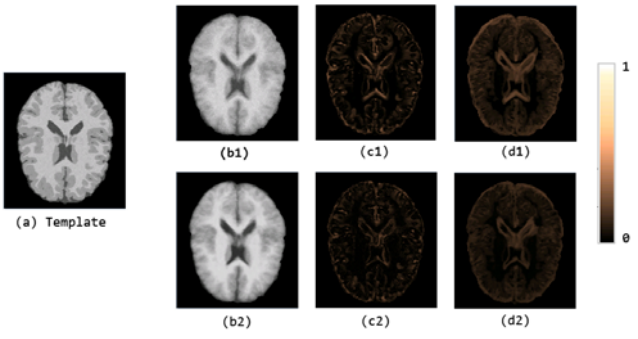


Fig. 1. (a)Template image. Top row: the mean image, the difference between the template and the mean image, the standard deviation image for the data-set before registration. Bottom row, the respective images for the group after registration.

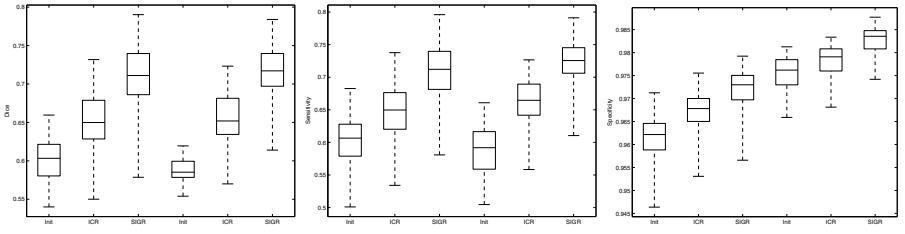


Fig. 2. From left to right, DICE coefficients, sensitivity and specificity for gray matter and white matter segmentations initially (Init) and after applying the iconic registration method (ICR) and the simultaneous geometric-iconic registration method (SIGR) respectively. The first three results are for the gray matter while the next ones for the white matter.

the iconic registration and are presented in Fig. 2. From the comparison of the DICE values, it can be concluded that the addition of the landmark information has ameliorated the registration result as the DICE values for the geometric - iconic registration are greater than the values of the iconic one. Moreover, it can be concluded that the addition of the landmark information has rendered the registration less sensitive to the initial conditions. This can be justified by the difference between the worst case results produced by both methods.

4 Discussion

In this paper we have proposed a novel approach to couple geometric (landmark) and iconic (voxel-wise) registration. The proposed method is, to the best of our knowledge, the first to propose an one-shot optimization on the joint parameter space, and therefore inherits ability to capture large deformations, independence with respect to the initial conditions, smooth and continuous diffeomorphic dense field, while being able to account for various similarity metrics and arbitrary number and position of landmarks. Promising results demonstrate the potentials of this elegant formulation.

The bias introduced from the landmark extraction process is an important limitation of the method. Such a limitation can be dealt with the use of the notion of missing correspondences. This is something that we are willing to address in the near future. The use of higher order model interactions between graph nodes is also interesting, since it could make the framework rigid or similarity invariant. Last but not least, the encouraging results that were obtained in the intra-modality case suggest that the application of the proposed method in the problem of multi-modal image fusion could be of significant interest.

References

1. Joshi, S., Miller, M.: Landmark matching via large deformation diffeomorphisms. In: IEEE TIP (2000)
2. Shen, D., Davatzikos, C.: Hammer: hierarchical attribute matching mechanism for elastic registration. In: IEEE TMI (2002)
3. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. In: CVIU (2003)
4. Beg, M.F., Miller, M.I., Trouvé, A., Younes, L.: Computing large deformation metric mappings via geodesic flows of diffeomorphisms. CVIU (2005)
5. 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)
6. Glocker, B., Komodakis, N., Tziritas, G., Navab, N., Paragios, N.: Dense image registration through mrfs and efficient linear programming. In: MedIA (2008)
7. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. NeuroImage (2009)
8. Johnson, H., Christensen, G.: Consistent landmark and intensity-based image registration. In: IEEE TMI (2002)
9. Cachier, P., Mangin, J.F., Pennec, X., Riviere, D., Papadopoulos-Orfanos, D., Règis, J.: Multisubject non-rigid registration of brain MRI using intensity and geometric features. In: Niessen, W.J., Viergever, M.A. (eds.) MICCAI 2001. LNCS, vol. 2208, p. 734. Springer, Heidelberg (2001)
10. Azar, A., Xu, C., Pennec, X., Ayache, N.: An interactive hybrid non-rigid registration framework for 3d medical images. In: IEEE ISBI (2006)
11. Biesdorf, A., Wörz, S., Kaiser, H.J., Stippich, C., Rohr, K.: Hybrid spline-based multimodal registration using local measures for joint entropy and mutual information. In: Yang, G.-Z., Hawkes, D., Rueckert, D., Noble, A., Taylor, C. (eds.) MICCAI 2009. LNCS, vol. 5761, pp. 607–615. Springer, Heidelberg (2009)
12. Hellier, P., Barillot, C.: Coupling dense and landmark-based approaches for nonrigid registration. In: IEEE TMI (2003)
13. Papademetris, X., Jakowski, A.P., Schultz, R.T., Staib, L.H., Duncan, J.S.: Integrated intensity and point-feature nonrigid registration. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 763–770. Springer, Heidelberg (2004)
14. Zhan, Y., Shen, D.: Deformable segmentation of 3-d ultrasound prostate images using statistical texture matching method. In: IEEE TMI (2006)
15. Ou, Y., Davatzikos, C.: Dramms: Deformable registration via attribute matching and mutual-saliency weighting. In: Prince, J.L., Pham, D.L., Myers, K.J. (eds.) IPMI 2009. LNCS, vol. 5636, pp. 50–62. Springer, Heidelberg (2009)
16. Kolmogorov, V.: Convergent tree-reweighted message passing for energy minimization. IEEE PAMI 28 (2006)

Groupwise Registration by Hierarchical Anatomical Correspondence Detection

Guorong Wu[1], Qian Wang[1,2], Hongjun Jia[1], and Dinggang Shen[1]

[1] Department of Radiology and BRIC, University of North Carolina at Chapel Hill
{grwu,jiahj,dgshen}@med.unc.edu
[2] Department of Computer Science, University of North Carolina at Chapel Hill
qianwang@cs.unc.edu

Abstract. We present a novel feature-based groupwise registration method to simultaneously warp the subjects towards the common space. Due to the complexity of the groupwise registration, we resort to decoupling it into two easy-to-solve tasks, i.e., alternatively establishing the robust correspondences across different subjects and interpolating the dense deformation fields based on the detected sparse correspondences. Specifically, several novel strategies are proposed in the correspondence detection step. *First*, attribute vector, instead of intensity only, is used as a morphological signature to guide the anatomical correspondence detection among all subjects. *Second*, we detect correspondence only on the driving voxels with distinctive attribute vectors for avoiding the ambiguity in detecting correspondences for non-distinctive voxels. *Third*, soft correspondence assignment (allowing for adaptive detection of multiple correspondences in each subject) is also presented to help establish reliable correspondences across all subjects, which is particularly necessary in the beginning of groupwise registration. Based on the sparse correspondences detected on the driving voxels of each subject, thin-plate splines (TPS) are then used to propagate the correspondences on the driving voxels to the entire brain image for estimating the dense transformation for each subject. By iteratively repeating correspondence detection and dense transformation estimation, all the subjects will be aligned onto a common space simultaneously. Our groupwise registration algorithm has been extensively evaluated by 18 elderly brains, 16 NIREP, and 40 LONI data. In all experiments, our algorithm achieves more robust and accurate registration results, compared to a groupwise registration method and a pairwise registration method, respectively.

1 Introduction

Registration of a population data has received more and more attention in recent years due to its importance in population analysis [1-4]. Since groupwise registration method is able to register all images without explicitly selecting the template, it can avoid bias in template selection and thus becomes attractive to the precise analysis of population data, compared to the pairwise registration methods. However, it is complicated for groupwise registration of multiple images simultaneously.

Although groupwise registration can be achieved by exhausting pairwise registrations between all possible subject combinations in the population [4], this type of

T. Jiang et al. (Eds.): MICCAI 2010, Part II, LNCS 6362, pp. 684–691, 2010.

method suffers from very heavy computation. Recently, more favorable approaches were proposed to align all subjects simultaneously by following the groupwise concept explicitly. Specifically, Joshi *et al.* [1] proposed to perform groupwise registration by iteratively (1) registering all subjects to the group mean image, and (2) constructing the group mean image as the Fréchet mean of all registered images. Also, Learned-Miller [3] proposed a congealing method to jointly warp the subjects towards a hidden common space by minimizing the sum of stack entropies in the population. Balci *et al.* [2] further extended the congealing method to non-rigid image registration by modeling the transformations with B-Splines. However, as the method itself is intensity-based, it is intrinsically insufficient to establish good anatomical correspondences across images. Furthermore, although the groupwise registration can be solved through steepest descent optimization [2, 3], it is unfortunately sensitive to local minima. Also, since the cost function is estimated based on only ~1% randomly sampled voxels (regardless of their morphological importance), the registration performance could be seriously affected.

To the best of our knowledge, the issue of anatomical correspondence in groupwise registration, which is very critical to measure the inter-subject difference, has not been well addressed in the literature. In this paper, we propose a novel feature-based groupwise registration method for achieving robust anatomical correspondence detection. Specifically, we formulate our groupwise registration by alternatively (1) estimating the sparse correspondences across all subjects and (2) interpolating the dense transformation field based on the established sparse correspondences.

In Step (1), we use attribute vector, instead of intensity only, as a morphological signature to help guide correspondence detection. Furthermore, the robustness of correspondence detection based on attribute vectors is achieved in two ways. *First*, we only detect correspondences for the most distinctive voxels, called as driving voxels, in the brain images, and then use their detected correspondences to guide the transformations of the nearby non-driving voxels. *Second*, multiple correspondences are allowed to alleviate the ambiguities particularly in the beginning of registration, and these one-to-many correspondences are gradually restricted to one-to-one correspondence with progress of registration in order to achieve accuracy for the final registration results. It is worth noting that this soft assignment strategy is also applied to all subjects in the population, where the contributions from different subjects are dynamically controlled through the registration. In Step (2), TPS is utilized to interpolate the dense transformation fields based on the sparse correspondences.

We have compared the performance of our groupwise registration with the congealing method [2] and the pairwise HAMMER registration algorithm [5, 6] by evaluation on 18 elderly brains, 16 NIREP dataset with 32 manually delineated ROIs, and 40 LONI dataset with 54 manually labeled ROIs. Experimental results show that our method can achieve the best performance.

2 Methods

Given a group of subjects $\boldsymbol{S} = \{S_i | i = 1 \dots N\}$, the goal of the groupwise registration is to find a set of transformations $G=\{g_i | g_i(x) = x + h_i(x),\ x = (x_1, x_2, x_3) \in \Re^3,\ i = 1, \cdots, N\}$ that are able to transform each subject towards a hidden common space

with its individual displacement $h_i(x)$. Fig. 1 schematically illustrates the concept of our groupwise registration. To connect each pair of subjects through the common space, the inverse transformation fields $G^{-1}=\{g_i^{-1}|g_i^{-1}(x) = x + h_i^{-1}(x),\ x = (x_1, x_2, x_3) \in \Re^3, i = 1, \cdots, N\}$ should be calculated as in Fig. 1. Thus the composite transformation $g_j^{-1} \circ g_i$ can be used to warp subject S_i to S_j, and also $g_i^{-1} \circ g_j$ can be used to warp subject S_j to S_i (as shown in the right panel of Fig. 1). In the following, we use i to index the subject under consideration and j for any other subject except S_i. Also, we call G the set of *forward transformations* and G^{-1} the set of *backward transformations*, which enter and leave the common space, respectively. In the following, we will present the energy function in Section 2.1, and then provide a solution to groupwise registration in Section 2.2.

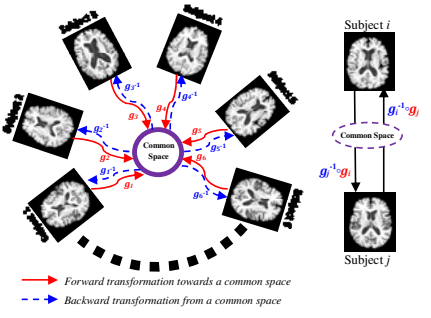


Fig. 1. The schematic illustration of the proposed groupwise registration algorithm. All subjects in the group are connected by the forward transformations g_i (i.e., red solid arrows) to the common space (i.e., a purple circled region), and by the backward transformations g_i^{-1} (i.e., blue dashed arrows) coming from the common space. The right panel shows the composite transformations bridging subjects S_i and S_j.

2.1 Energy Function in Groupwise Registration

As pointed in Fig. 1, all subjects will agglomerate to the hidden common space by following the simultaneously estimated transformation fields. To identify the anatomical correspondence among all subjects, we propose using attribute vector as a morphological signature of each voxel x for guiding the correspondence detection. Without loss of generality, the geometric moment invariants of white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) are calculated from a neighborhood of each voxel x for defining its attribute vector $\vec{a}(x)$ [5]. Further, we hierarchically select distinctive voxels (with distinctive attribute vectors) as driving voxels in the image [5], by adaptively setting thresholds on the attribute vectors. Here, the driving voxels are represented as $\{x_{i,m}|i = 1, \ldots, N, m = 1, \ldots, M_i\}$, where M_i is the number of the driving voxels in subject S_i. In our groupwise registration, we establish the sparse correspondences only on the driving voxels due to their distinctiveness, and let these driving voxels steer the dense transformation, by considering them as control points in TPS-based interpolation [7].

The overall energy function of groupwise registration can be defined to minimize the differences of attribute vectors on the corresponding locations across different subjects and also preserve the smoothness of estimated transformation fields:

$$E(G) = \sum_{i=1}^{N} \sum_{x \in S_i} \sum_{j=1, j \neq i}^{N} \left\| \vec{a}_i(x) - \vec{a}_j\left(g_j^{-1} \circ g_i(x)\right) \right\|^2 + \sum_{i=1}^{N} \|Lg_i\|^2, \quad (1)$$

where L is an operator to compute the bending energy of transformation field g_i.

However, directly optimizing $E(G)$ is usually intractable. Thus, we introduce the sparse correspondence fields $F = \{f_i(x) | x \in \Re^3, i = 1, \ldots, N\}$ in our method. Here, each f_i is a set of correspondence vectors defined for subject S_i: for each driving voxel $x_{i,m}$ in subject S_i, it gives the latest estimated corresponding location (pointing to the common space), while for each non-driving voxel, it keeps the previous estimated transformation. As the result, the energy function in Eq. 1 becomes:

$$E(F, G) = \sum_{i=1}^{N} \sum_{x \in S_i} \left\{ \sum_{j=1, j \neq i}^{N} \left\| \vec{a}_i(x) - \vec{a}_j\left(g_j^{-1} \circ f_i(x)\right) \right\|^2 + \|f_i(x) - g_i(x)\|^2 \right\} + \sum_{i=1}^{N} \|Lg_i\|^2, \quad (2)$$

The advantage of introducing F is that it decouples the complicated optimization problem into two simple-to-solve sub-problems, i.e., alternatively (**SP$_1$**): estimating the correspondence field f_i via correspondence detection; and (**SP$_2$**): interpolating the dense transformation g_i with regularization on G.

Estimating the Correspondence Field F (SP$_1$): In this step, we take advantage of the driving voxels to establish the correspondence on each driving voxel $x_{i,m}$ of subject S_i by inspecting each candidate in a neighborhood n_1, w.r.t. each of other subjects S_j one by one. For evaluating each candidate, several useful strategies are employed here to achieve robust correspondences. *First*, not only the voxelwise but also the regionwise difference on attribute vectors is proposed by computing the distance of each pair of corresponding attribute vectors within a neighborhood n_2. *Second*, multiple spatial correspondences are allowed on each driving voxel $x_{i,m}$ by introducing a spatial assignment $\pi_{i,m}^{j,v}$ to indicate the likelihood of the true correspondence v w.r.t. subject S_j. Also, we use $\tau_{i,m}^{j}$ to describe the likelihood of subject S_j being selected as a reference image for correspondence detection of $x_{i,m}$ of subject S_i, according to the similarity of local morphology between S_i and S_j.

Therefore, by fixing G in Eq. 2, the new energy function E_1 in this step can be defined as:

$$\begin{aligned} E_1(\pi, \tau, F) &= \sum_{i=1}^{N} \sum_{m=1}^{M_i} E_{corr}(\pi_{i,m}, \tau_{i,m}, f_i) + \sum_{i=1}^{N} \sum_{m=1}^{M_i} E_{fuzzy}(\pi_{i,m}, \tau_{i,m}) + \sum_{i=1}^{N} \sum_{m=1}^{M_i} \left\| f_i(x_{i,m}) - g_i(x_{i,m}) \right\|^2 \\ E_{corr}(\pi_{i,m}, \tau_{i,m}, f_i) &= \sum_{j=1}^{N} \tau_{i,m}^{j} \sum_{v \in n_1\left(g_i(x_{i,m})\right)} \pi_{i,m}^{j,v} \cdot \left[\left\| g_i(x_{i,m}) - v \right\|^2 + d\left(\vec{a}_i(x_{i,m}), f_i, g_j^{-1}\right) \right] \\ E_{fuzzy}(\pi_{i,m}, \tau_{i,m}) &= r_1 \cdot \pi_{i,m}^{j,v} \cdot log \pi_{i,m}^{j,v} + r_2 \cdot \tau_{i,m}^{j} \cdot log \tau_{i,m}^{j} \end{aligned} \quad (3)$$

where $d\left(\vec{a}_i(x_{i,m}), f_i, g_j^{-1}\right) = \sum_{z \in n_2(x_{i,m})} \left\| \vec{a}(z) - \vec{a}\left(g_j^{-1}(f_i(z))\right) \right\|^2$ measures the regionwise difference of attribute vector $\vec{a}_i(x_{i,m})$ and its corresponding counterpart in S_j w.r.t. the

current estimated correspondence f_i and the previous obtained inverse transformation field g_j^{-1}. There are totally three terms in the energy function E_1. The first term E_{corr} measures the matching discrepancy on each driving voxel $x_{i,m}$, where the criteria in evaluating the candidate v w.r.t. subject S_j are: 1) the spatial distance between $g_i(x_{i,m})$ and v in the common space should be as small as possible according to the ICP principle [8]; 2) not only the candidate location v but also its neighborhood should have the similar attribute vectors based on the measurement $d(\vec{a}_i(x_{i,m}), f_i, g_j^{-1})$.

Soft assignment is very important for brain image registration to reduce the risk of mismatching, particularly in the beginning of registration. All voxels in the search neighborhood n_1 have the chance to become correspondence candidate, but their contributions to the true correspondence vary according to the matching criteria. To increase the registration accuracy and specificity, it is also necessary to evolve to one-to-one correspondence in the end of registration. Therefore, the second term $E_{fuzzy}(\pi_{i,m}, \tau_{i,m})$ is used to dynamically control the soft assignment by requiring the entropy of $\pi_{i,m}$ and $\tau_{i,m}$ gradually to decrease with progress of registration.

The third term in E_1 ensures that the correspondence field F be close to the previous estimated transformation field G, by minimizing the difference between each pairs of f_i and g_i.

Interpolating the Dense Transformation Field G (SP_2): After updating the correspondence in each driving voxel, the energy function in this step is given as:

$$E_2(G) = \sum_{i=1}^{N} \sum_{m=1}^{M_i} \left\{ \left\| f_i(x_{i,m}) - g_i(x_{i,m}) \right\|^2 \right\} + \sum_{i=1}^{N} \|Lg_i\|^2. \tag{4}$$

By regarding the driving voxels as control points in each subject, TPS interpolation can be used to estimate the optimal g_i that fits the transformation on $x_{i,m}$ to $f_i(x_{i,m})$ and reaches the minimal bending energy (the second term) [7,9].

2.2 Implementation for Groupwise Registration

In our method, we alternatively optimize SP_1 and SP_2 in each round of registration. In SP_1, the explicit solutions of $\pi_{i,m}^{j,v}$ and $\tau_{i,m}^{j}$ are obtained by letting $\partial E_1 / \partial \pi_{i,m}^{j,v} = 0$ and $\partial E_1 / \partial \tau_{i,m}^{j} = 0$:

$$\pi_{i,m}^{j,v} = c_1 \cdot exp\left(-\frac{\xi_v}{r_1}\right), \qquad \pi_{i,m}^{j,v} \leftarrow \pi_{i,m}^{j,v} \Big/ \sum_{v \in n_2(g_i(x_{i,m}))} \pi_{i,m}^{j,v} \tag{5}$$

$$\tau_{i,m}^{j} = c_2 \cdot exp\left(-\frac{\sum_{v \in n_2(g_i(x_{i,m}))} \pi_{i,m}^{j,v} \xi_v}{r_2}\right), \quad \tau_{i,m}^{j} \leftarrow \tau_{i,m}^{j} \Big/ \sum_{j=1}^{N} \tau_{i,m}^{j} \tag{6}$$

where $\xi_v = \left[\left\| g_i(x_{i,m}) - v \right\|^2 + d(\vec{a}_i(x_{i,m}), f_i, g_j^{-1}) \right]$ denotes the bracket part in $E_{corr}(\pi_{i,m}, \tau_{i,m}, f_i)$, and c_1 and c_2 are constants. After $\pi_{i,m}^{j,v}$ and $\tau_{i,m}^{j}$ are determined by Eqs. 5 and 6 on each $x_{i,m}$, the correspondence $f_i(x_{i,m})$ is updated by moving it to the mean location of all candidates under the guidance of $\pi_{i,m}^{j,v} \cdot \tau_{i,m}^{j}$:

$$\hat{f}_i(x_{i,m}) = \sum_{j=1}^{N} \sum_{v \in n_2(g_i(x_{i,m}))} \tau_{i,m}^{j} \cdot \pi_{i,m}^{j,v} \cdot v. \quad (7)$$

It is worth noting that we introduce a *multi-resolution strategy* to implement our proposed groupwise registration method for fast and robust registration. Specifically, in each resolution, the size of the search neighborhood n_1 decreases gradually with the progress of registration, for achieving more specific detection of correspondences. Moreover, in the initial state of registration, only a small set of voxels with distinctive features, such as those locating at ventricular boundaries, sulcal roots and gyral crowns, are selected as driving voxels. After that, more and more driving voxels are added to drive the registration and eventually all voxels in the brains join the groupwise registration.

By taking the driving voxels as control points, TPS is used to interpolate the dense transformation field after we repeat the calculations of Eqs. 5~7 for each driving voxel $x_{i,m}$. To avoid the cumbersome inversion of large matrix in TPS (proportional to the number of control points), we perform TPS interpolation in overlapping blocks ($32 \times 32 \times 32$) and also down-sample the driving voxels in each block.

3 Experiments

In our experiments, we have extensively evaluated the performances of our groupwise registration method in atlas building and ROI labeling. For comparison, we use the congealing groupwise registration method [2] with its available codes, http://www.insight-journal.org/browse/publication/173. To demonstrate the advantage of groupwise registration over pairwise registration, the registration results by a pairwise registration method, namely HAMMER [5], are also provided.

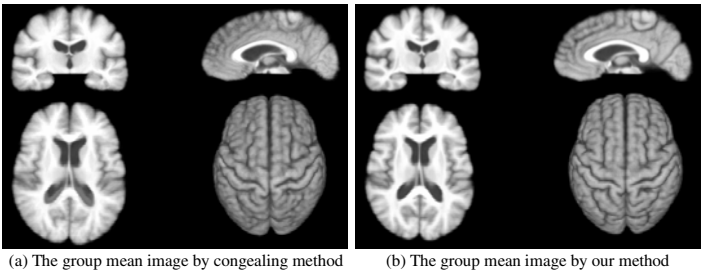
To demonstrate the group overlap of labeled brain regions after registration, we specifically vote a reference by assigning each voxel with a tissue label that is the majority of all tissue labels at the same location from all aligned subjects. Then, the overlap ratio between each of the registered label images and the voted reference can be calculated. Here, we use the Jaccard Coefficient metric as the overlap ratio to measure the alignment of the two regions (A and B) with the same label, defined as:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}. \quad (8)$$

18 Elderly Brain Images: 18 elderly brain images, each with $256 \times 256 \times 124$ voxels and the resolution of $0.9375 \times 0.9375 \times 1.5mm^3$, are used in this experiment. The group mean images produced by the congealing method and our groupwise registration method are both displayed in Fig. 2. Through visual inspection, the group mean of our method is sharper and gains better contrast (especially around ventricles) than that of the congealing method, indicating better performance of our registration method. The overlap ratios on WM, GM, and VN, as well as the overall overlap ratio on the whole brain, by our method and the congealing method, are provided in Table 1. It can be observed that our method achieves better results than the congealing method in each tissue type. On the other hand, it is interesting to compare the performance between groupwise and pairwise registrations of these 18 brain images. In

HAMMER-based pairwise registration, 5 out of 18 subjects are randomly selected as the templates to align all other 17 remaining subjects. The average overlap ratios produced by these five different templates, as well as the standard deviations, are shown in the first row of Table 1, which verify again the power of our groupwise registration in consistently registering the population data.

NIREP Data and LONI Data: In this experiment, we employ pairwise HAMMER, congealing method, and our groupwise registration method to align 16 NIREP data (with 32 manual ROIs) and LONI40 dataset (with 54 manual ROIs). Table 2 shows the average overlap ratios on these two datasets by the three registration methods. Obviously, our groupwise registration method achieves the most accurate registration results among all three registration methods. In particular, Fig. 3 shows the performance of registration accuracy at the left and the right precentral gyri of 16 NIREP brains by the two groupwise methods. The average overlap ratio is 45.95% by congealing and 57.34% by our method. The brighter color indicates the higher consistency of registration across different subjects, while the darker color means the poor alignment. Again, our method achieves much better alignment.



(a) The group mean image by congealing method (b) The group mean image by our method

Fig. 2. The groupwise registration results by the congealing method and our method. It can be observed that our group mean image is much sharper than that by the congealing method, indicating a more accurate and consistent registration by our method.

Table 1. Overall overlap ratios of WM, GM, and VN by pairwise HAMMER algorithm, congealing method, and our groupwise registration method

	WM	GM	VN	Overall
Pairwise HAMMER	63.86% (±3.87%)	57.25% (±2.18%)	76.51% (±3.70%)	65.64% (±3.15%)
Congealing Method	59.68%	51.09%	70.61%	59.43%
Our Method	75.81%	63.61%	81.16%	73.52%

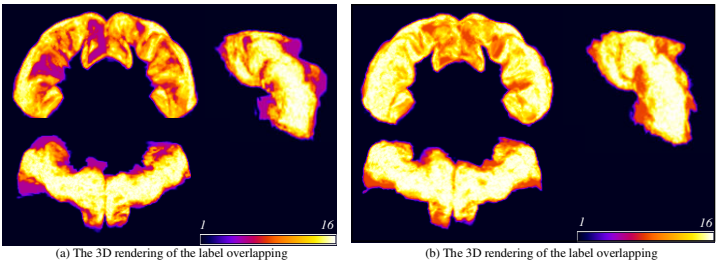
4 Conclusion

In this paper, we have presented a new feature-guided groupwise registration method and also demonstrated its applications in atlas building and population data analysis. Specifically, by taking advantage of the driving voxels (with distinctive features) automatically detected from all images, we develop a feature-based groupwise registration method by alternatively estimating the correspondences on the driving voxels and updating the dense transformation fields by TPS. Extensive experiments have

been performed to compare the performance of our method with that of the congealing method and the pairwise HAMMER algorithm. All experimental results show that our method can achieve the best performance.

Table 2. Overall overlap ratios of the aligned ROIs in NIREP and LONI datasets by pairwise HAMMER algorithm, congealing method, and our groupwise registraion method

	Pairwise HAMMER	Congealing Method	Our Method
NIREP (32 ROIs)	56.58%	52.07%	61.52%
LONI (54 ROIs)	54.12%	60.60%	67.02%



(a) The 3D rendering of the label overlapping results by the congealing method

(b) The 3D rendering of the label overlapping results by our groupwise registration method

Fig. 3. 3D renderings of the aligned left and right precentral gyri by the congealing method and our groupwise registration method

References

1. Joshi, S., Davis, B., Jomier, M., Gerig, G.: Unbiased diffeomorphic atlas construction for computational anatomy. NeuroImage 23, 151–160 (2004)
2. Balci, S.K., Golland, P., Shenton, M., Wells, W.M.: Free-Form B-spline Deformation Model for Groupwise Registration. In: Workshop on Open-Source and Open-Data for 10th MICCAI, pp. 105–121 (2007)
3. Learned-Miller, E.G.: Data driven image models through continuous joint alignment. IEEE Transactions on Pattern Analysis and Machine Intelligence 28, 236–250 (2006)
4. Shattuck, D.W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K.L., Poldrack, R.A., Bilder, R.M., Toga, A.W.: Construction of a 3D probabilistic atlas of human cortical structures. NeuroImage 39(3), 1064–1080 (2008)
5. Shen, D., Davatzikos, C.: HAMMER: Hierarchical attribute matching mechanism for elastic registration. IEEE Transactions on Medical Imaging 21(11), 1421–1439 (2002)
6. Shen, D., Davatzikos, C.: Very high resolution morphometry using mass-preserving deformations and HAMMER elastic registration. NeuroImage 18(1), 28–41 (2003)
7. Bookstein, F.L.: Principal Warps: Thin-Plate Splines and the Decomposition of Deformations. IEEE Transactions on Pattern Analysi and Machine Intelligence 11(6), 567–585 (1989)
8. 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)
9. Chui, H., Rangarajan, A.: A new point matching algorithm for non-rigid registration. Computer Vision and Image Understanding 89, 114–141 (2003)

Author Index